An Efficient Polynomial Chaos-based Proxy Model for History Matching and Uncertainty Quantification of Complex Geological Structures

Hamid Bazargan

Submitted for the
Degree of Doctor of Philosophy
Institute of Petroleum Engineering
Heriot-Watt University
May 2014

Abstract

A novel polynomial chaos proxy-based history matching and uncertainty quantification method is presented that can be employed for complex geological structures in inverse problems. For complex geological structures, when there are many unknown geological parameters with highly nonlinear correlations, typically more than 10^6 full reservoir simulation runs might be required to accurately probe the posterior probability space given the production history of reservoir. This is not practical for high-resolution geological models. One solution is to use a "proxy model" that replicates the simulation model for selected input parameters. The main advantage of the polynomial chaos proxy compared to other proxy models and response surfaces is that it is generally applicable and converges systematically as the order of the expansion increases. The Cameron and Martin theorem 2.24 states that the convergence rate of the standard polynomial chaos expansions is exponential for Gaussian random variables. To improve the convergence rate for non-Gaussian random variables, the generalized polynomial chaos is implemented that uses an Askey-scheme to choose the optimal basis for polynomial chaos expansions [199]. Additionally, for the non-Gaussian distributions that can be effectively approximated by a mixture of Gaussian distributions, we use the mixture-modeling based clustering approach where under each cluster the polynomial chaos proxy converges exponentially fast and the overall posterior distribution can be estimated more efficiently using different polynomial chaos proxies.

The main disadvantage of the polynomial chaos proxy is that for high-dimensional problems, the number of the polynomial chaos terms increases drastically as the order of the polynomial chaos expansions increases. Although different non-intrusive methods have been developed in the literature to address this issue, still a large number of simulation runs is required to compute high-order terms of the polynomial chaos expansions. This work resolves this issue by proposing the reduced-terms polynomial chaos expansion which preserves only the relevant terms in the polynomial chaos representation. We

demonstrated that the sparsity pattern in the polynomial chaos expansion, when used with the Karhunen-Loéve decomposition method or kernel PCA, can be systematically captured.

A probabilistic framework based on the polynomial chaos proxy is also suggested in the context of the Bayesian model selection to study the plausibility of different geological interpretations of the sedimentary environments. The proposed surrogate-accelerated Bayesian inverse analysis can be coherently used in practical reservoir optimization workflows and uncertainty assessments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The world's economic growth is estimated to stay around 3% per year until at least 2030 [33]. Consequently, it will drive global demand for energy. It is estimated that the global energy demand growth to average 1.2% per year from 2013 to 2030 [33]. The composition of the world's energy has evolved over time, influenced by a range of factors entailing cost, availability and technology. Figure 1.1 illustrates that in 1900, coal was by far the most dominant source of energy while seventy years later, oil surpass coal to become the largest energy source with more than 40 percent and natural gas usage raised to almost 20 percent [33]. By 2000, while oil maintained to be the main source of the global energy, natural gas grew at similar levels to coal and newer sources of energy appears in the mix [33]. Oil and natural gas combined formed nearly 60 percent of the total global energy usage in 2005 [33]. By 2030, oil will remain the main source of energy supply in the world at an estimate of 34 percent while natural gas will grow to reach 25 percent of the global energy usage [33]. Accordingly, oil and gas will have to provide more than 60 percent of the world's energy demand by 2030. ExxonMobil's assessment on the energy outlook demonstrates that the current production rate cannot meet this ever increasing demand and new production plans are required [33]. The optimum production plan needs a comprehensive knowledge about the reservoir performance prediction.

The conventional methods to predict reservoir performance is decline curve analysis and material balance [37]. Decline curve analysis fits a graphical curve to the observed production data and uses the curve to infer future performance. The decline curve requires a start point, usually taken right after the plateau time, a curvature and a rate of decline. In practice, the logarithm of production rate is plotted against the cumulative production and based on a mathematical model (harmonic, hyperbolic or

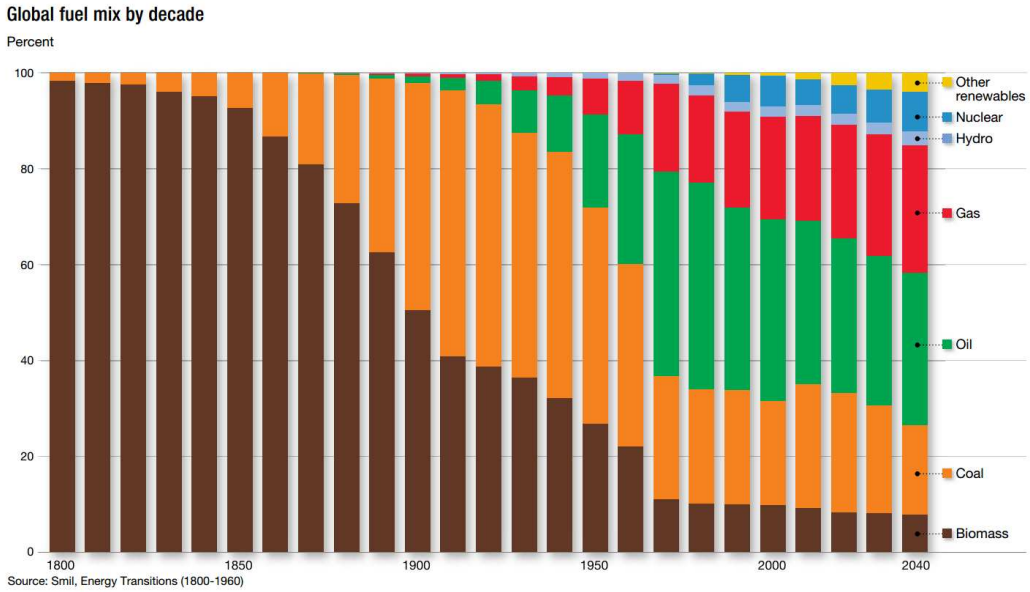

Figure 1.1: Evolving energy resources [33].

exponential) a curve is fitted to the plot. As Decline curve analysis does not involve any physical interpretation of subsurface flow, it fails to predict reservoir performance when the development strategy is changed. Additionally no pressure data is involved in the analysis and only under certain circumstances the future production rate can be inferred [37].

Material balance is based on the assumption that the reservoir volume is constant and the depletion of a reservoir volumetric component is compensated by the expansion of other components and the aquifer influx. A simple description of material balance analysis is given in Dake [37]:

$$\begin{aligned} \text{Withdrawal} = {} & \text{expansion of oil and solution gas left in the reservoir} \\ & + \text{expansion of the gas cap} + \\ & + \text{ reduction in pore volume due to rock compressibility} \\ & + \text{ reduction in hydrocarbon pore volume due to connate water expansion} \\ & + \text{ aquifer influx.} \end{aligned} \tag{1.1}$$

From the volumetric analysis and the characteristics of fluids and rock (PVT tests), the relationship between the fluid expansion and the average reservoir pressure can be deduced. The main drawback of material balance analysis is the assumption that the reservoir is homogenous and also reacts spontaneously and equally throughout its

entire volume. Consequently, not enough information about the possible development scenarios (i.e., the location of new wells) can be inferred using material balance analysis.

A more complex and robust method to predict reservoir performance is reservoir simulation technique. It involves discretization of reservoir into grid blocks, each of which has a different set of locally averaged properties of rock and fluids that dictate flow. Accordingly, It uses material balance analysis on each grid block to calculate a numerical approximation of flow rates of reservoir fluids from one block to the adjacent blocks. By combining material balance equations for all the grid blocks of reservoir and applying boundary conditions, pressure and flow rates of reservoir fluids at each grid block can be approximated by solving a system of equations.

The microscopic study of reservoir fluids flow at each grid block is infeasible as the distribution of pore spaces in rock is not known. In practice, the size of a grid block can be as large as a high rise building, e.g. $20m \times 20m \times 10m$, hence the microscopic study has to be replaced with a macroscopic approach. A mathematical model commonly used in reservoir simulation assigns some static parameters to each grid block and use the *Darcy* equation [38] to approximate the average velocity of fluids in and out of each grid block. The key static parameters used by the mathematical model for each grid block are the porosity (ϕ), absolute permeability ($\mathbf{k}$), relative permeability ($\mathbf{k_r}$), fluid saturations ($\mathbf{S_o}, \mathbf{S_w}, \mathbf{S_g}$), the dimension of the block, depth and pressure.

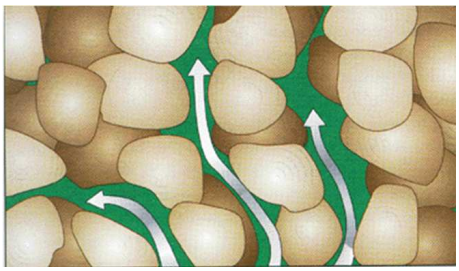

Figure 1.2: The microscopic study of flow through porous media is not practical as the distribution of pore spaces in rock is impossible to model.

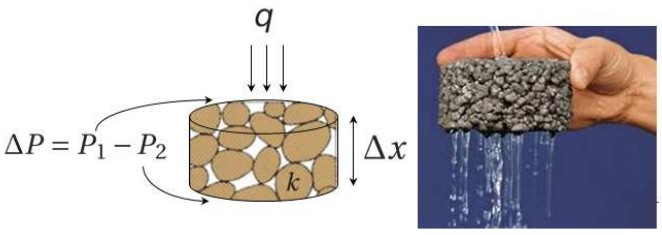

Figure 1.3: The Darcy equation uses the permeability concept to calculate the volumetric flow rate q through a porous medium.

The Darcy equation [38] originates in the work of Henry Darcy, who conducted an experiment on optimal filter designs for the water supply of Dijon, from an active aquifer. For a single phase fluid through a porous medium, the Darcy equation uses

the permeability value ($\mathbf{k}$) to calculate the volumetric flow rate q;

$$q = \frac{\mathbf{k}A}{\mu}\frac{\Delta P}{\Delta x} \tag{1.2}$$

where ΔP is the pressure difference over distance Δx, μ is the viscosity of fluid and A is the cross-sectional area that fluid is flowing through. The average velocity of fluid can be accordingly written as:

$$\overline{u} = \frac{\mathbf{k}}{\mu}\frac{\Delta P}{\Delta x}. \tag{1.3}$$

In the limit, when Δx is sufficiently small, the partial differential form of equation (1.3) can be derived as

$$u = \frac{\mathbf{k}}{\mu}\frac{\partial P}{\partial x}. \tag{1.4}$$

The effect of gravity can be taken into account by simply adding the corresponding pressure gradient in the direction of flow, to the term $\frac{\partial P}{\partial x}$.

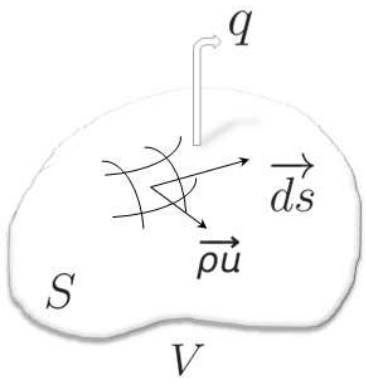

Figure 1.4: Mass conservation for a single phase flow through an arbitrary volume V enclosed by surface S.

For a single phase fluid flow through a porous medium, if we write the mass balance equation for any arbitrary volume V enclosed by surface S, we have:

$$\frac{d}{dt}\left(\int_V \rho dv\right) = -\oint_S \rho\overrightarrow{u}.\overrightarrow{ds} + \int_V q dV \tag{1.5}$$

where $\rho\overrightarrow{u}$ is the transport influx through $\overrightarrow{ds}$ and q is the sink/source term. The negative sign for the term $(-\iint_S \rho\overrightarrow{u}.\overrightarrow{ds})$ is present because a positive outward flux corresponds to a negative rate of mass change. Using the Divergence theorem [90]:

$$\oint_S \rho\overrightarrow{u}.\overrightarrow{ds} = \int_V \nabla.(\rho u) dv. \tag{1.6}$$

Consequently for any arbitrary volume V we have:

$$\begin{aligned} &\int_V \big(\frac{dP}{dt} + \nabla.(\rho u)\big)dv = \int_V q dV \\ &\Rightarrow \nabla.(\rho u) + \frac{dP}{dt} = q \end{aligned} \tag{1.7}$$

For multi-phase fluid flow, the interaction of reservoir fluids must be incorporated into the flow equations. Fluids are either miscible or immiscible. Miscible fluids mix perfectly together to form a solution, as in the case of oil and the gas dissolved. Immiscible fluid do not mix together at any portion, as in the case of oil and water. The interaction of miscible fluids affects the fluids viscosity and can be directly incorporated into the flow equations. The interaction of immiscible fluids are modeled using relative permeability curves ($\mathbf{k_r}$). The curve is plotted against the fluid saturation and models the drop in permeability due to the presence of other fluids. For two-phases system the relative permeability curve is measured experimentally, but for three-phases the interactions of each pair of fluids have to be incorporated accordingly [171]. Consequently, the relative permeability curves can be used to calculate the effective permeability for each phase $\mathbf{k_e}$ in the equation (1.4), where $\mathbf{k_e} = \mathbf{k_r} \cdot \mathbf{k}$. For example, the Darcy velocity of water can be written as:

$$u_w = \frac{\mathbf{k_e}}{\mu_w}\frac{\partial P_w}{\partial x} = \frac{\mathbf{k_r}(\mathbf{s_w})\mathbf{k}}{\mu_w}\frac{\partial P_w}{\partial x} \tag{1.8}$$

where $\mathbf{s_w}$ is the water saturation and $\frac{\partial P_w}{\partial x}$ is the gradient of water pressure. The difference between the pressure of the existing fluid phases can be modeled using capillary pressure curves [53].

Phase behavior in reservoirs is either addressed by the black oil model or a compositional model. In the black oil model, oil, water and gas are treated as a separate phases and the fluid properties are modeled using empirical equations based on laboratory experiments [183]. Compositional model treats the fluid as a mixture of different hydrocarbon molecules, water and other non-carbon based components of reservoir fluid [32].

For a multi-phase fluid flow in a reservoir, writing a mass balance equation for each component in each phase separately, results in a complex set of partial differential equations for each grid block. The system of partial differential equations can be numerically solved for all the reservoir grid blocks, with respect to boundary conditions, and for discrete points in time called *time steps*. Reservoir simulation is the science of using computer programs to solve a system of complex differential equations that governs flow and transport in natural porous geological formations.

Reservoir simulation techniques have been steadily progressing during the past decades with the advances of computing hardware. In the early 1950s, reservoir simulators could handle only a two dimensional reservoir model with simple geometry, while in the 1960s simulators progressed to model a three dimensional realistic geometry of reservoirs and incorporated the black oil model to simulators to handle phase behavior [191]. Compositional models were introduced in the 1970s as an alternative to the black oil model to simulate enhanced oily recovery and gas injection processes [191]. In the 1980s, reservoir simulators progressed to handle complex well managements and fractured reservoir models [191]. During the 1990s, reservoir visualization enhanced by developing graphical used interfaces [191]. By 2000, reservoir simulators were provided with features such as local grid refinement, unstructured grids and the ability to integrate with surface facilities [191]. Figure 1.5 illustrates that reservoir simulations evolved along similar paths with microprocessors computing capabilities during over the past four decades [159].

By 2004, frequency scaling technique nearly reached its limit because of the increasing power consumption necessary to achieve higher frequency and consequently higher speed for microprocessors [164]. Hence, hardware designers turned to contrive multicore processors. The use of parallel processing revolutionized reservoir simulators with the number of grids that could be accommodated. Parallel processing works on the principle that large problems can be effectively broken into smaller ones that are then solved simultaneously on separate microprocessors as clusters of a single machine connected by a network [11]. Accordingly, the dominant force to improve computational performance was shifted from frequency scaling technique that increases the speed of microprocessors in serial processing to parallel solvers that efficiently breaks large problems into smaller ones. Then, the main hallmark of current reservoir simulators is the use of the parallel processing technique, which can simulate reservoirs with more than 10^{10} grid blocks in a reasonable time.

One common practice to estimate reservoir model variables is to use observed reservoir behavior, so called *History Matching*. History matching is a type of inverse problem in which instead of solving the forward problem for a set of reservoir model variables to assess reservoir performance, observed production history of reservoir is used to estimate reservoir model parameters. History matching problems are almost always ill-posed and non unique where different sets of reservoir model variables may result in equally good matches with observed history of reservoir [126]. Consequently, a single history matched reservoir model is not sufficient to predict reservoir performance and an assessment of uncertainty in reservoir model variables, and accordingly in reservoir per-

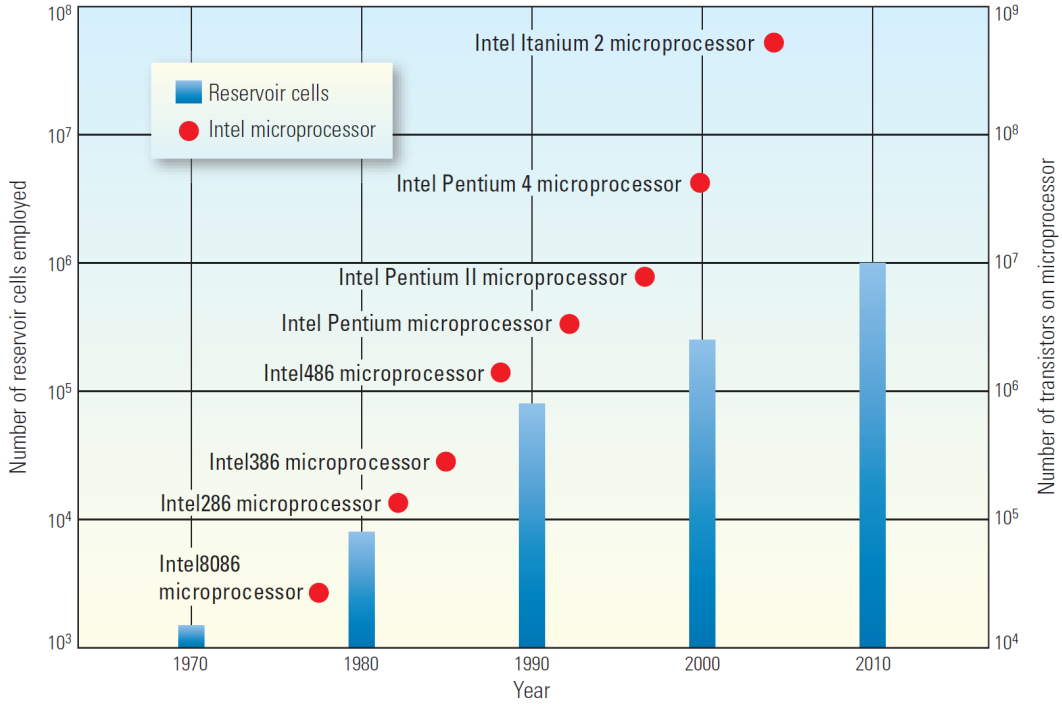

Figure 1.5: The evolution of computing capability versus reservoir simulation [159]. The shift of computing architecture from simply increasing the number of transistors (higher frequency) to the parallel processing, has improved reservoir simulators with the number of grid blocks that could be accommodated.

formance prediction, is the required part of the solution to a history matching problem. Albeit, obtaining even one single history-matched reservoir model is computationally very expensive and usually requires thousands of reservoir simulation runs [126]. However, the ability to generate multiple reservoir models has remarkably progressed in the past decades, mainly because of the substantial growth in computational capability and also the use of *Monte Carlo* methods and *geostatistics* [126].

Methods of history matching have been reviewed a number of times in the literature [200, 113, 52, 190, 126]. In the groundwater context, Yeh [200] classified the techniques for model parameter estimation as either direct or indirect and thoroughly reviewed the existing methods under each category. McLaughlin and Townley [113] reviewed the methods of functional analysis for groundwater model calibration problem. In the petroleum literature, Ewing et al. [52] and Watson et al. [190] reviewed the gradient-based techniques for the history matching problem and discussed about different parameterization methods to mitigate the ill-posedness of the problem. Solving a history matching problem as a minimization problem requires an objective function that measures the mismatch with observed history of reservoir. Oliver et al. [126]

presented a comprehensive literature review of history matching techniques used for different forms of objective function.

Methods for uncertainty quantification can be generally classified into:

1. methods that compute a single best model but use covariance or sensitivity matrix to quantify uncertainty

2. methods that draw multiple samples from the posterior distribution.

Floris et al. [55] presented a comparative study of uncertainty quantification methods for history matching problem of a synthetic reservoir model (PUNQ-S3). Barker et al. [12] updated the study by applying importance sampling and Markov Chain Monte Carlo (MCMC) to the PUNQ-S3 synthetic reservoir model. They concluded that the adaptive framework of MCMC gives rigorous results while it is only several times more expensive than the other methods.

If the history matching problem is solved in the Bayesian context, with a prior distribution for reservoir model variables and a likelihood probability density function given the observed history of reservoir, then MCMC methods are the most comprehensive tool to generate samples from the posterior distribution of reservoir model variables. In the reservoir simulation context, MCMC was first applied by Oliver et al. [128] to history matching of noisy transient pressure. They concluded that the acceptance rate of MCMC to generate uncorrelated samples from the posterior was too low to be practical for history matching and uncertainty quantification of real fields. Several variants of MCMC method have been examined in the context of history matching to achieve higher acceptance rate but still the computational cost is a limiting factor [126]. For example, Bonet-cunha et al. [18] used hybrid MCMC that combines a stochastic step with a deterministic step for the proposal scheme to improve the acceptance rate of the Markov chain. However, the computational cost was too high to be of practical application [126].

Although the growth in computation capabilities has advanced numerical simulation in assisted history matching, for many workflows in uncertainty quantification the computational cost is still a major limiting factor. For example, a typical use of Markov Chain Monte Carlo method requires to run more than 10^6 full reservoir simulations which is not feasible. A "proxy model" or "surrogate model" is a mathematical or statistical model that replicates the simulation-model for selected input parameters. Several successful applications of different proxy models in production optimization, reservoir history matching and sensitivity analysis of uncertain reservoir model variables have been reported in literature [54, 131, 69, 10, 130, 201, 137, 99, 87, 168]. Zubarev et al.

[203] presented a comparative study of proxy-modeling methods and investigated the predictive quality and computational efficiency of different proxy modeling algorithm. They concluded that all proxy-modeling techniques showed a strong dependence on the model complexity, dimension of the problem, and quality of the trial runs [203]. Hence the optimum choice of proxy is problem specific and any decision based on the proxy models requires a thorough understanding of their limitations. Polynomial chaos proxy has this significant advantage over other proxy models that for all applications it guarantees the convergence in probability to the output random variable of interest, i.e. cumulative oil production. Convergence of polynomial chaos proxy to the reservoir simulation-model happens in probability, and also in distribution, as the order of polynomial chaos expansion increases. The use of polynomial chaos expansion for uncertainty quantification was pioneered by Ghanem and Spanos [64] and has been applied to various engineering problems. In reservoir simulation context, Sarma et al. [152] used polynomial chaos expansion to efficiently quantify uncertainty for closed-loop production optimization and model updating. This work aims mainly toward developing a robust workflow for history matching and uncertainty quantification using polynomial chaos proxy. The objectives of the thesis are as follows:

- To construct an efficient and also generally applicable proxy model based on polynomial chaos expansion, capable of continual model updating of the complex geological structures
- To develop an efficient sampling framework based on polynomial chaos proxy using Markov Chain Monte Carlo
- Obtaining an analytical approximation for the posterior probability distribution of reservoir model variables using polynomial chaos expansion
- To specify the effective cross terms in polynomial chaos representation for the reduced dimension parameters of reservoir model
- To improve the convergence rate of polynomial chaos proxy by using mixture-model based clustering of reservoir model variables
- To establish a probabilistic framework for comparing different scenarios emanated from different interpretation of sedimentary environments
- To integrate multiple reservoir descriptions using alternative training images and also sample from the posterior distribution of reservoir model parameters for each training image.

The thesis is structured as follows:

Chapter 2 presents the Wiener polynomial chaos representation of random variables with finite second moments and reviews the convergence properties of polynomial chaos expansions under Cameron and Martin theorem. It also studies different techniques to compute the coefficients of polynomial chaos expansions followed by presenting the generalized polynomial chaos expansions, which uses other types of orthogonal polynomials to improve the rate of convergence for non-Gaussian random variables.

Chapter 3 proposes a history matching work flow based on the polynomial chaos expansion which incorporates the semi-definite programming to find the global minimum of the misfit surface. We conduct a two dimensional synthetic case study of fluvial channels to demonstrate the efficiency of the approach.

Chapter 4 applies polynomial chaos proxy in conjunction with the MCMC method to efficiently sample from the posterior probability density function of reservoir model parameters. We present a two dimensional example of fluvial channels to demonstrate that the high-order polynomial chaos proxy can capture the effects of remarkably nonlinear forward model on the posterior distribution of reservoir model parameters.

Chapter 5 uses clustering approach to study the "scenario uncertainty" where the prior reservoir model has a high level of discrete uncertainty. We study the flow response uncertainty under each scenario and integrate the uncertainties to achieve the overall uncertainty. Our approach to cluster analysis in mixture modeling is based on the posterior probabilities of models, known literally as "Bayesian Model Selection". We apply the proposed cluster-based polynomial chaos proxy framework to study the plausibility of three training images based on different geological interpretation of the second layer of synthetic Stanford VI reservoir, based on the given data.

Chapter 6 presents the concluding remarks in addition to some insights to future work for the application of polynomial chaos proxy in history matching and uncertainty quantification.

Chapter 2

Polynomial Chaos Expansion

2.1 Introduction

The uncertainties can be characterized as either aleatory uncertainties, which are irreducible uncertainties inherent in nature, or epistemic uncertainties, which are reducible uncertainties associated with a lack of knowledge [93]. The epistemic uncertainties can be attenuated with the better knowledge of the system, process or mechanism. However, to quantify aleatory uncertainty, probabilistic methods are commonly used when sufficient information is available. To quantify aleatory uncertainty, a fundamental step is the accurate representation of random phenomenon using a reasonable number of degrees of freedom. A popular approach, known by the names polynomial chaos expansion, Fourier-Hermite expansion or Wiener-Hermite expansion, represents a random variable by a series of Hermite polynomials in a countable sequence of independent Gaussian random variables and uses truncations of such expansions as approximations. These random variables are called "basic random variables".

The root of Wiener's "polynomial chaos expansion", can be found in a series of papers from the period 1919-1922, of which "Differential Space" [194] was the penultimate, and in the later paper "Generalized Harmonic Analysis" [195]. However, the term polynomial chaos was originally introduced in his 1938 paper [196], in which he applies his generalized harmonic analysis [195] to a mathematical formulation of statistical mechanics. He used the Brownian motion as the motivation. In that work, Wiener introduced the concept of a continuous homogeneous chaos, which in modern statistical terminology corresponds approximately to a homogeneous random field defined on $\mathbb{R}^d$. Wiener extended his theory on "pure one-dimensional chaos", which was

basically a random measure defined by the increments of the Wiener process, to the Wiener homogenous chaos expansion. Subsequently, Cameron and Martin [25] proved that any square-integrable functional (with respect to Wiener measure) can be represented as a series of Hermite polynomials in a countable sequence of Gaussian random variables which converges in the mean square sense. The connection between multiple Wiener integrals and Fourier Hermite expansion is accordingly derived at Itô's work [77]. The Wiener's contribution on polynomial chaos is also described in [109]. A modern exposition of Hermite expansions of functionals of Brownian motion can be found in [72, 79, 88].

Throughout the Wiener's work, there is emphasis on the models of natural phenomena, particularly on models of Brownian motion. He expressed random quantities as functions of Brownian motion. In mathematics, Brownian motion $B(t)$ is defined as a continuous-time stochastic process which has the following properties:

- $B(0) = 0$ with probability 1
- $B(t)$ is almost surely continuous
- $B(t)$ has independent increments with distribution of disjoint intervals: $B(t + dt) - B(t) \sim \mathcal{N}(0, dt)$

which $\mathcal{N}(0, dt)$ denotes the Gaussian distribution with the expected value of zero and variance of dt. A sample path can be thought of as the definite integral, from time zero, of the sample of the Gaussian distribution that constitutes the history of buffering to which the Brownian particle was subject. This is indeed the basis on which Wiener [195] defined the Wiener measure. Wiener tried to model a stochastic variable X as functionals of Brownian motions, $X = g(B)$. In the literature of stochastic analysis there are three basic representations for square-integrable functionals of Brownian motion:

- polynomial chaos expansions
- stochastic Itô integrals
- multiple Wiener integrals.

There exist deep connections between these representations and each can be converted to the others. However, the work of Ghanem and Spanos [64] on stochastic finite element methods, where random processes representing input and solutions of stochastic differential equations were approximated with a truncated Hermite polynomials of Gaussian random variables, sparked renewed interest in the polynomial chaos expansions. Their

work was a pivotal contribution in this context as they used the truncated polynomial chaos expansions in a Galerkin framework and proposed the spectral stochastic finite elements methods. Renewed interest in the polynomial chaos expansions was also due to the recent developments in computational methods for solving stochastic partial differential equations [8, 9, 110, 181, 198].

Often, the truncated polynomial chaos approximations for X, denoted by X_n, are used for calculations. In this case, it becomes necessary to study the convergence properties of a sequence of random variables $\{X_n, n \geqslant 1\}$ approximating X. It is well-known that the polynomial chaos approximations for second order random variables (random variables with finite variance) converges in the mean-square sense [64, 88]. Further, this condition implies that the sequence of the polynomial chaos approximations $\{X_n, n \geqslant 1\}$ converges to X in probability and in distribution. To illustrate the convergence properties of two sequence of random variables, in the next section we review the different modes of convergence, as well as relationships between them.

This chapter is structured as follows: after reviewing the convergence modes at Section 2.2, Section 2.3 enumerates the representation of the standard Wiener-Hermite polynomial chaos expansions, including the Cameron-Martin theorem. Section 2.4 treats the various techniques to compute the coefficients of polynomial chaos expansion, entailing a discussion on the various technical issues. Section 2.5 presents the generalized polynomial chaos expansions, which uses other types of orthogonal polynomials, instead of Hermite polynomials, to improve the rate of convergence for non-Gaussian random processes.

2.2 Convergence Concepts

Let X and $X_n, n \geqslant 1$, be real-valued random variables defined on probability space (Ω,$\mathcal{F}$,P) and the distribution functions of X and X_n are F and F_n, respectively. A probability space consists of three parts [108]:

- A sample space, Ω, which is the set of all possible outcomes.
- A set of events $\mathcal{F}$, where each event contains a set of outcomes. An outcome is the result of a single execution of the model.
- The assignment of probabilities to the events using the probability measure function, P.

To facilitate the discussion on the convergence properties of the polynomial chaos expansions, in this section we review different convergence modes as well as relationships between them.

2.2.1 Modes of Convergence

Depending on the way in which the difference between the sequence of random variables $\{X_n, n \geqslant 1\}$ and X is measured, different definitions for the convergence can be specified. These definitions are commonly referred as modes of convergence.

- **Almost Surely Convergence**

  The sequence of $X_n, n \geqslant 1$, is said to converge almost surely (a.s) or with probability 1 to X, written $X_n \xrightarrow{a.s} X$, if

  $$\lim_{n\to\infty} X_n(\omega) = X(\omega), \qquad \forall \omega \in \Omega \setminus A, \tag{2.1}$$

  where $P(A) = 0$.
  Convergence with probability 1 is the strongest mode of convergence for a sequence of random variables.

- **Convergence in probability**
  The sequence of $X_n, n \geqslant 1$, converges in probability (i.p), written $X_n \xrightarrow{i.p} X$ if for all $\epsilon > 0$,

  $$\lim_{n\to\infty} P(|X - X_n| > \epsilon) = 0. \tag{2.2}$$

The concept of convergence in probability is frequently used in statistics. For example, an estimator is called "consistent" if it converges to the quantity of interest in probability. The convergence in probability means that the possibility of an "unusual" outcome becomes smaller as the sequence progresses. The *continuous mapping theorem* states that for every continuous function $g(.)$, if $X_n \xrightarrow{i.p} X$ then also $g(X_n) \xrightarrow{i.p} g(X)$ [13]. For example, if the sequence of X_n converges in probability to X, hence the sequence e^{X_n} converges toward e^X in probability. It is also worth noting that convergence in probability defines a topology on the space of random variables over a fixed probability space.

Definition 1. *Let $F(X) = P(X \leq x)$, $x \in \mathbb{R}$ denote the distribution function of random variable X; we say $X \in L_p$, $p > 0$, if*

$$E[|X|^p] = \int_{-\inf}^{\inf} |X|^p dF(x) < \infty. \tag{2.3}$$

Note that for special cases, $p = 1$ and $p = 2$, $E[|X|] < \infty$ and $E[|X|^2] < \infty$ respectively, meaning that X has a finite mean and finite variance.

- **Convergence in L_p Sense**

The sequence of $X_n, n \geqslant 1$, is said to converge in L_p sense or in the p-th mean to X, written $X_n \xrightarrow{L_p} X$, if $\forall p > 1$, $X \in L_P$ and

$$\lim_{n\to\infty} E[|X_n - X|^p] = 0. \tag{2.4}$$

We refer to the special case of $p = 1$, *convergence in mean* and for $p = 2$, *the mean-square convergence.*

- **Convergence in Distribution**

The sequence of $X_n, n \geqslant 1$ with probability distribution of $F_n(X)$ is said to converge in distribution to X, written $X_n \xrightarrow{d} X$, if

$$\lim_{n\to\infty} F_n(X) = F(X), \qquad X \in \mathbb{R}. \tag{2.5}$$

The convergence in distribution heuristically states that if n is sufficiently large,

the probability of X_n to be in a given range is approximately equal to the probability of X to be in that range. Convergence in distribution is also referred to as weak convergence [144]; because, as opposed to convergence in probability or L_p convergence, convergence in distribution only probes the behavior of the probability distribution of the random variables.

2.2.2 Connection between Modes of Convergence

Almost sure convergence is the strongest mode of convergence and it implies convergence in probability. Therefore, if two sequence of random variables is almost surely convergent, they also converge in probability as well. Also, convergence in probability implies convergence in distribution [184]. Accordingly, almost sure convergence implies convergence in distribution. The chain of implication between these convergence modes are as follows:

$$X_n \xrightarrow{a.s} X \Longrightarrow X_n \xrightarrow{i.p} X \Longrightarrow X_n \xrightarrow{d} X. \tag{2.6}$$

Convergence in $L_p, p \geq 1$ implies convergence in probability. This can be shown by making use of Markov's inequality [170] which states that for a random variable X, if $\epsilon > 0$ and $p \geq 1$, then we have:

$$P(|X| \geq \epsilon) \leq E|X|^p/\epsilon^p. \tag{2.7}$$

Hence, it is straightforward to see convergence in L_p leads to convergence in probability and accordingly it deduces to convergence in distribution. The chain of implication between these convergence modes are as follows:

$$X_n \xrightarrow{L_p} X \Longrightarrow X_n \xrightarrow{i.p} X \Longrightarrow X_n \xrightarrow{d} X. \tag{2.8}$$

It also states that the higher order convergence in L_p leads to convergence in the lower order [184]. Therefore if two random variables are convergent in L_r and $r > s$, they are also convergent in L_s norm.

2.3 Wiener-Hermite Polynomial Chaos Expansions

In this section, we review the standard Wiener polynomial chaos expansions with independent normal distributions as the basic random variables. Then we present the

Cameron-Martin theorem [25] which studies the convergence properties of the Wiener polynomial chaos expansions, followed by few illustrative examples.

2.3.1 Representation of Wiener Polynomial Chaos Expansions

Wiener [196] represented a general second-order random variable $X(\theta)$, viewed as a function of $\theta \in \Omega$ (the independent random event), in the following form:

$$\begin{aligned} X(\theta) = a_0 H_0 + \sum_{i_1=1}^{\infty} a_{i_1} H_1(\xi_{i_1}(\theta)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1,i_2} H_2(\xi_{i_1}(\theta), \xi_{i_2}(\theta)) \\ + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1,i_2,i_3} H_3(\xi_{i_1}(\theta), \xi_{i_2}(\theta), \xi_{i_3}(\theta)) + \dots . \end{aligned} \tag{2.9}$$

where $H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ denotes the orthogonal Hermite polynomials of order n and $(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ are multi-dimensional independent Gaussian random variables with zero mean and unit variance. The equation (2.9) is the discrete version of the original Wiener polynomial chaos expansion introduced in his 1938 paper [196], where the continuous integrals are replaced by summations. The Hermite polynomials of order n, $H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ can be derived by the following formulation [196]:

$$H_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n}) = e^{(\frac{1}{2}\xi^T \xi)} (-1)^n \frac{\partial^n}{\partial \xi_{i_1} \partial \xi_{i_2} \dots \partial \xi_{i_n}} e^{-(\frac{1}{2}\xi^T \xi)}. \tag{2.10}$$

Here, ξ denotes the vector of n Gaussian random components.

The one dimensional Hermite polynomials can be derived using the equation (2.10), $H_n(\xi) = e^{\frac{1}{2}\xi^2}(-1)^n \frac{\partial^n}{\partial \xi^n} e^{-\frac{1}{2}\xi^2}$, as:

$$\begin{aligned} H_0(\xi) &= 1 \\ H_1(\xi) &= \xi \\ H_2(\xi) &= \xi^2 - 1 \\ H_3(\xi) &= \xi^3 - 3\xi. \end{aligned} \tag{2.11}$$

The higher orders ($n > 2$) can be alternatively obtained using the following relationship:

$$H_n(\xi) = \xi H_{n-1}(\xi) - (n-1) H_{n-2}(\xi). \tag{2.12}$$

The Hermite polynomials form an orthogonal basis in the *Hilbert* space [146] of functions on $\mathbb{R}$ in which the inner product is given by the integral including the Gaussian weight function $w(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$;

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(\xi) g(\xi) w(\xi) d\xi. \tag{2.13}$$

Under the Gaussian probability measure ($w(\xi) : \mathcal{N}(0,1)$), Hermite polynomials are orthogonal to each other and form a complete basis of Hilbert space. In other words, Wiener's idea was to project any functionals of Brownian motion in the Hilbert space constructed by Hermite polynomials. In Figure 2.1, the one dimensional Hermite polynomials of different orders are depicted. The orthogonality of Hermite polynomials

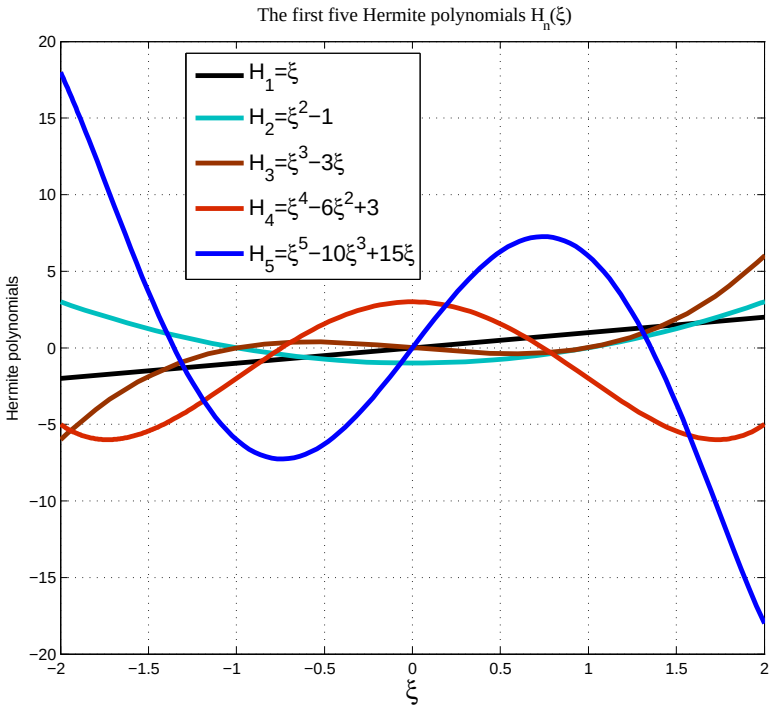

Figure 2.1: Hermite polynomials of different orders

under the Gaussian measure implies:

$$\langle H_i(\xi), H_j(\xi)\rangle = \int_{-\infty}^{\infty} H_i(\xi)H_j(\xi)e^{-\frac{\xi^2}{2\sigma^2}}d\xi = 0 \qquad \forall i,j \geq 1 \qquad i \neq j, \tag{2.14}$$

and also,

$$\langle H_i(\xi), H_i(\xi)\rangle = \int_{-\infty}^{\infty} H_i(\xi)H_i(\xi)\frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2\sigma^2}}d\xi = i! \qquad \forall i \geq 1 \tag{2.15}$$

For multi-dimensional Hermite polynomials, a simple way to construct the polynomials of order n with k number of random variables is to tensorize the one-dimensional Hermite polynomial, rather than direct use of the equation (2.10). The expression

for $H_n(\xi_1, \xi_2, \ldots, \xi_k)$ can be achieved by constructing the tensor-products of univariate polynomials. For example, the 3$^{\text{rd}}$-order of Hermite polynomials with two random variables $H_3(\xi_1, \xi_2)$ can be written as:

$$H_3(\xi_1, \xi_2) = \begin{cases} \{3,0\} & \Rightarrow H_3(\xi_1)H_0(\xi_2) = \xi_1^3 - 3\xi_1, \\ \{2,1\} & \Rightarrow H_2(\xi_1)H_1(\xi_2) = (\xi_1^2 - 1)\xi_2, \\ \{1,2\} & \Rightarrow H_1(\xi_1)H_2(\xi_2) = \xi_1(\xi_2^2 - 1), \\ \{0,3\} & \Rightarrow H_0(\xi_1)H_3(\xi_2) = \xi_2^3 - 3\xi_2. \end{cases} \tag{2.16}$$

Consequently, $H_3(\xi_1, \xi_2)$ in the equation (2.10), can be written as:

$$H_3(\xi_1, \xi_2) = a_{3,0}(\xi_1^3 - 3\xi_1) + a_{2,1}(\xi_1^2 - 1)\xi_2 + a_{1,2}\xi_1(\xi_2^2 - 1) + a_{0,3}(\xi_2^3 - 3\xi_2). \tag{2.17}$$

All higher orders of multi-dimensional Hermite polynomial way can be constructed similarly. The orthogonality condition of the equation (2.14) also holds for multi-dimensional Hermite polynomials, under the multi-dimensional Gaussian measure $w(\xi)$:

$$w(\xi) = \frac{1}{\sqrt{(2\pi)^{i_k}}} e^{-\frac{1}{2}\xi^T\xi} \tag{2.18}$$

To find the corresponding coefficient of each term in standard polynomial chaos representation (2.9), the orthogonality condition implies:

$$a_n = \frac{\langle H_n, X \rangle}{\langle H_n, H_n \rangle} = \frac{\int_{-\infty}^{\infty} H_n(\xi) X(\theta) w(\xi) d\xi}{\int_{-\infty}^{\infty} H_n^2(\xi) w(\xi) d\xi}. \tag{2.19}$$

In order to conduct the above projection, we need to transform the stochastic process $X(\theta)$ into the probability space with Gaussian measure. If X is already a function of independent Gaussian random variables, we can take them as the basic random variables and no probability space transformation is needed. Otherwise, both X and the basic random variables need to be transformed in the same probability space. Under the theory of probability this is always possible [132]. One idea is to map them to the uniformly distributed probability space $u \in U(0,1)$. Without loss of generality, we present the procedure when X a one-dimensional continuous random variable. Let us assume that the random variable u is uniformly distributed in $(0,1)$ and the probability density functions for X and ξ are $f(x)$ and $w(\xi)$, respectively. Since the cumulative distribution of X and ξ lies between 0 and 1, they can be transformed to the same uniformly distributed random variable u by:

$$u = \int_{-\infty}^{X} f(x)dx = \int_{-\infty}^{\xi} w(\xi)d\xi, \tag{2.20}$$

where $w(\xi)$ is the Gaussian measure. Now X and ξ have effectively mapped into probability space with uniform distribution (u) measure. Hence we can do the integration of the equation (2.19) by substituting

$$X = F^{-1}(u) \qquad and \qquad \xi = W^{-1}(u), \tag{2.21}$$

where F and W are their corresponding cumulative distribution function. We shall review an example of mapping into the Gaussian probability measure at the end of this section.

For multi-dimensional random variables, the multiple integration of the equation (2.19) may be cumbersome. To circumvent the direct integration, various techniques have been studied in the literature which will be discussed in Section 2.4.

2.3.2 Convergence properties of Wiener Polynomial Chaos Expansions

In applications of Wiener-Hermite polynomial chaos expansions the underlying Gaussian Hilbert space is often taken to be the space spanned by a given finite sequence of $\xi_m \in \mathbb{N}$ of independent Gaussian random variables. The polynomial chaos approximation can be achieved by truncating the expansion (2.9) into the p^{th} order of Hermite polynomials. Let $PC_p(\xi)$ denote the p-term polynomial chaos approximation of random variable $X(\xi)$:

$$PC_p(\xi) = \sum_{i=1}^{p} \alpha_i H_i(\xi), \tag{2.22}$$

where $H_i(\xi)$ is the Hermite polynomial of order i. Cameron and Martin [25] studied the convergence behavior of sequence $\{PC_1(\xi), PC_2(\xi), \ldots, PC_n(\xi)\}$ to X as $n \longrightarrow \infty$. In the following we present the Cameron-Martin theorem for convergence of the Hermite polynomial chaos representation of any functional $X : \mathbb{R} \to \mathbb{R}$ of a Gaussian random variable $\xi : \Omega \to \mathbb{R}$. The theory can be coherently extended to the multidimensional polynomial chaos expansions [48].

Cameron-Martin Theorem. *The Hermite chaos expansion of any second order functional $X(\xi)$ under Gaussian measure $w(\xi)$ converges in the mean-square sense to X. This implies that if X is a second-order stochastic process, i.e.,*

$$\int_{-\infty}^{\infty} |X(\xi)|^2 w(\xi) d\xi \quad < \quad \infty, \tag{2.23}$$

then

$$\int_{-\infty}^{\infty} |X(\xi) - PC_P(\xi)|^2 w(\xi) d\xi \quad \longrightarrow 0 \qquad as \qquad n \longrightarrow \infty, \tag{2.24}$$

where $PC_P(\xi)$ is the polynomial chaos approximation of order p given by the equation (2.22).

Following the discussion in Section 2.2, convergence in the mean-square proves convergence in mean (L_1), convergence in probability and convergence in distribution accordingly;

- **Polynomial chaos convergence in probability**
  Cameron-Martin Theorem implies that the sequence of $PC_p(\xi)$ converges in probability to $X(\xi)$. Hence, for any $\epsilon > 0$, $\exists\, n$ that for $p > n$, $P(|X(\xi) - PC_p(\xi)|) < \epsilon$. This means

$$\lim_{p \to \infty} P(|X(\xi) - PC_p(\xi)| > \epsilon) = 0. \tag{2.25}$$

- **Polynomial chaos convergence in distribution**
  Cameron-Martin Theorem implies that the sequence of $PC_p(\xi)$ converges in distribution to $X(\xi)$, i.e.,

$$\lim_{p \to \infty} P[PC_p(\xi) \leq x] = P[X(\xi) \leq x]. \tag{2.26}$$

  This property suggests that the quantiles of the sequence $PC_p(\xi)$ converge for $p \longrightarrow \infty$ to the corresponding quantiles of $X(\xi)$. (These can be set-valued).

Cameron and Martin [25] studied the rate of convergence for the standard Wiener-Hermite polynomial chaos approximation and demonstrated that the convergence rate is faster than exponential. Further, they have also shown that the error in the expansion decays as $\mathcal{O}(\frac{1}{(p+1)!})$, where p is the highest order Hermite polynomial used in the polynomial chaos approximation. For one dimensional PC expansion, they derived the following inequality for the convergence rate:

$$|X(\xi) - PC_p(\xi)| \leq \frac{C}{(p+1)!} \left| \frac{\mathbf{d}^{p+1} X(\xi)}{\mathbf{d}\xi^{p+1}} \right|, \tag{2.27}$$

where C is a constant. The above inequality can also be extended to the case when multidimensional Hermite polynomial are used as the basis functions [25].

It is worth noting that the convergence in the mean-square sense does not guarantee the convergence in L_m for $m > 2$. In the sense that $E[|X(\xi) - PC_p(\xi)|^m]$ does not necessarily go to 0 when $p \longrightarrow \infty$. Furthermore, convergence of polynomial chaos

expansion in the mean-square sense does not imply almost surely convergence of $PC_p(\xi)$ to $X(\xi)$.

Here we study two simple examples to illustrate the basic idea behind the standard Wiener-Hermite polynomial chaos expansion, in which the integration of the equation (2.19) is accessible. In the first example, there is no need for probability space transformation, but in the second example the probability space needs to be measured by the Gaussian probability to perform the integration. For high-dimensional problems, direct integration of the equation (2.19) is not practically feasible and other techniques must be used to compute the integral. We shall discuss these methods in detail in Section 2.4.

Example 2.3.1. ***The polynomial chaos expansion vs Taylor Series***

Let $g(\xi) = e^{\xi}$ where ξ is a standard Gaussian distribution $\xi \sim \mathcal{N}(0,1)$. Here, $g(\xi)$ is already a functional of Gaussian random variable ξ, hence we can take ξ as the measure with the weighting function $w(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$. To obtain the polynomial chaos approximation of $g(\xi)$, we need to compute the integral of the equation (2.19) to find the corresponding coefficient of one dimensional Hermite polynomial of order n 2.9:

$$g(\xi) \approx \alpha_0 H_0 + \sum_{i=1}^{n} \alpha_i H_i(\xi), \tag{2.28}$$

where $H_i(\xi)$ is the Hermite polynomial of order i, and α_i can be calculated as:

$$\alpha_i = \frac{\langle g(\xi), H_i \rangle}{\langle H_i, H_i \rangle} = \frac{\int_{-\infty}^{\infty} g(\xi) H_i(\xi) w(\xi) d\xi}{i!} = \frac{1}{\sqrt{2\pi} i!} \int_{-\infty}^{\infty} e^{\xi - \frac{\xi^2}{2}} H_i(\xi) d\xi = \frac{e^{\frac{1}{2}}}{i!}. \tag{2.29}$$

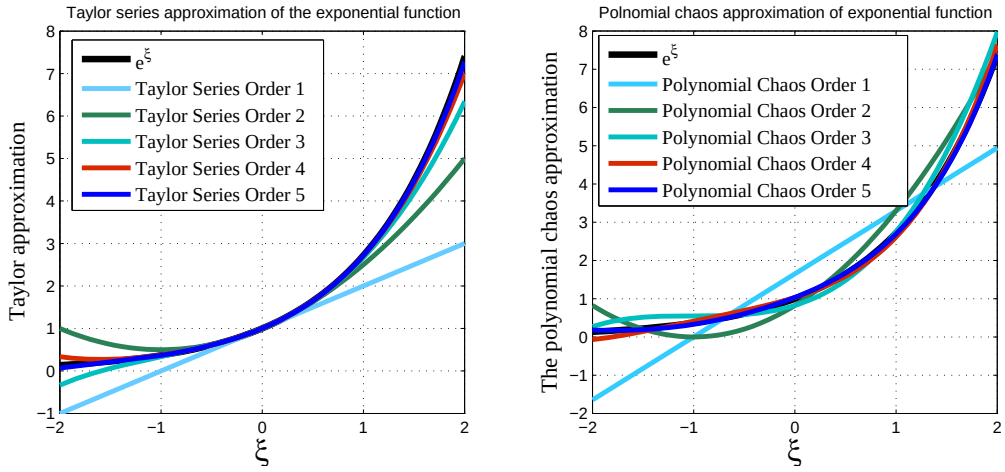

Figure 2.2: The polynomial chaos approximation (right) vs Taylor series (left) in the small neighborhood of 0, $[-2, 2]$, Example 2.3.1

Therefore, the equation (2.28) of the polynomial chaos approximation of order n for $g(\xi)$ *can be written as:*

$$g_{PC}^{n}(\xi) = \sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi) = e^{\frac{1}{2}}(1 + \xi + \frac{(\xi^2 - 1)}{2!} + \frac{(\xi^3 - 3\xi)}{3!} + \ldots + \frac{H_n(\xi)}{n!}). \quad (2.30)$$

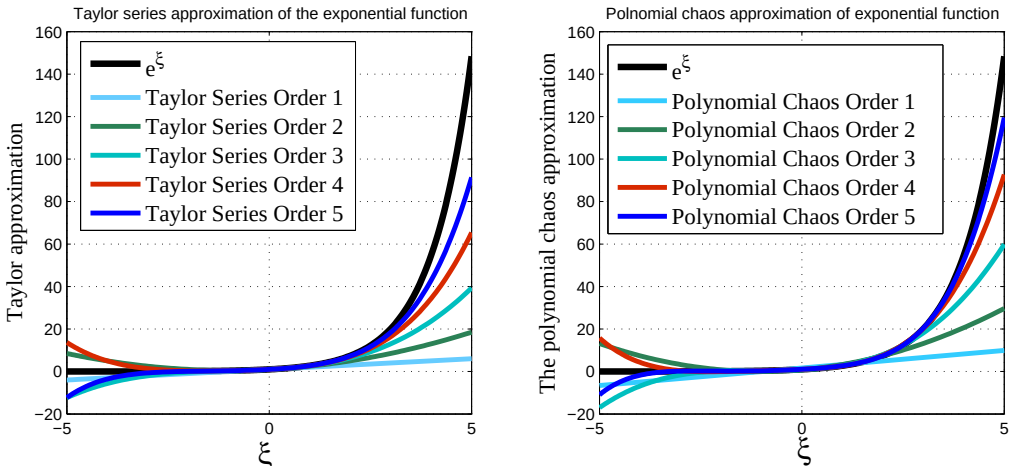

Figure 2.3: The polynomial chaos approximation (right) vs Taylor series (left) in the wider range $[-5, 5]$, Example 2.3.1

It is instructive to compare the polynomial chaos approximation of the equation (2.30) to the corresponding Taylor series of the exponential function;

$$g_{T}^{n}(\xi) = 1 + \sum_{i=1}^{n} \frac{\xi^i}{i!} = 1 + \xi + \frac{\xi^2}{2!} + \frac{\xi^3}{3} + \frac{\xi^4}{4!} + \ldots + \frac{\xi^n}{n!} \quad (2.31)$$

in which Taylor series is written in the neighborhood of 0*. Figure 2.2 compares the Taylor series vs the polynomial chaos approximation, in the small neighborhood of* 0*, and Figure 2.3 brings it to a wider range.*

In a small range, Taylor series is slightly more accurate than the polynomial chaos approximation, while in the wider range the polynomial chaos expansion is more precise. Now the question is, which of them is a more comprehensive tool for uncertainty quantification purpose?

To answer the above question, we have to take into account the fact that not every value for ξ *happens with the same probability. For example the probability of* ξ *being zero is* $P(\xi = 0) = 0.3989$*, while the probability of* $P(\xi = 1) = 0.242$*. Hence it is reasonable to define a weighted error by aggregating the approximation error,* $error(\xi) = (g(\xi) - \hat{g})^2$*, with the Gaussian weighting function* $w(\xi) \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$*. Figure 2.4 shows the weighted*

error of Taylor series compared to the polynomial chaos approximation. As the order of polynomial chaos expansion increases the precision of the approximation enhances progressively and the weighted error tends to go to zero with an exponential rate (The semi-log plot of Figure 2.5 illustrates that the mean-square error of the polynomial chaos approximation decays with an exponential rate as the order of the expansion increases).

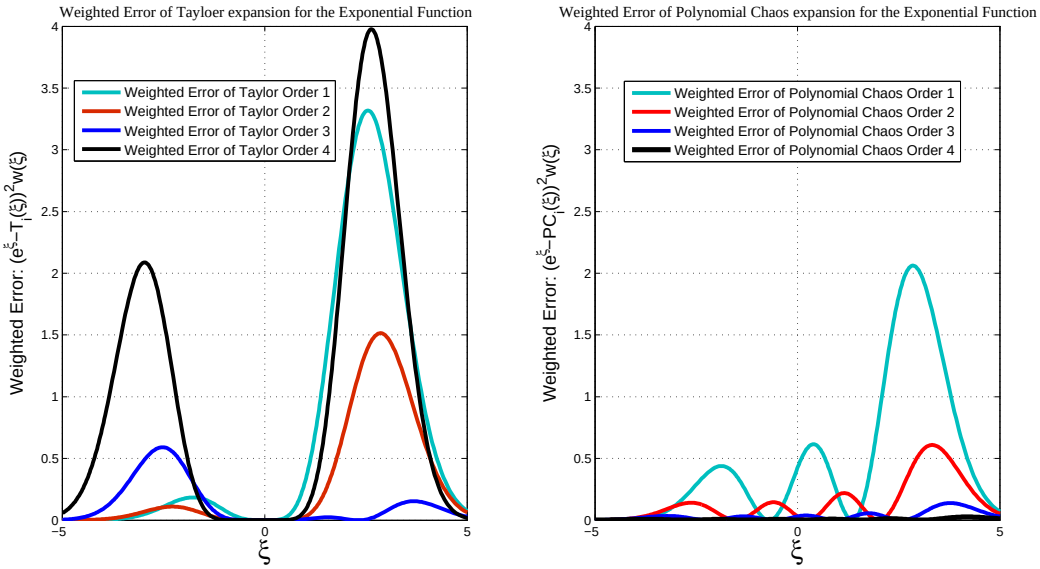

Figure 2.4: The weighted error of approximations by the polynomial chaos expansions (right) and Taylor series (left) for the exponential function of Example 2.3.1

Figure 2.4 suggests that the polynomial chaos approximation exhibits much lower weighted error than Taylor series approximation. In statistics, the Mean-Square Error (MSE) is the most important criterion used to evaluate the performance of an estimator and it is defined by

$$MSE(\xi) \;=\; E(g(\xi)-\hat{g})^2 \;=\; \int_{-\infty}^{\infty}(g(\xi)-\hat{g})^2 e^{-\frac{\xi^2}{2}}d\xi. \tag{2.32}$$

Table 2.1 compares the MSE of the polynomial chaos approximation with Taylor series up to 5^{th}-order. It demonstrates that mean-square error of the polynomial chaos expansion vanishes faster than Taylor series. Hence, it is a more accurate tool to represent the exponential function $g(\xi) = e^{\xi}$ when ξ is a Gaussian random variable, and consequently more appropriate for uncertainty quantification.

Table 2.1: Mean-square error of the polynomial chaos approximation vs Taylor series

	Taylor Series	Polynomial Chaos
First Order	279.4171	195.2492
Second Order	124.6728	59.3352
Third Order	46.5100	14.0305
4^{th} Order	14.8457	2.7043
5^{th} Order	4.1303	0.4390

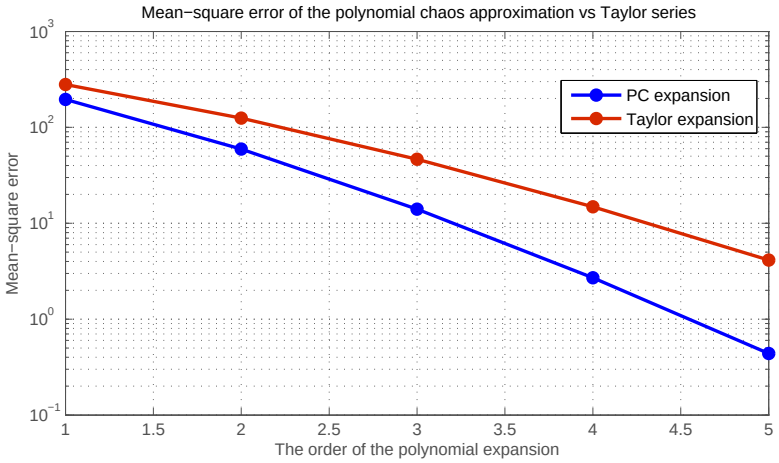

Figure 2.5: The mean-square error of the polynomial chaos approximation vs Taylor series for the exponential function of Example 2.3.1

Example 2.3.2. ***Gaussian Mixture Model***

In probability and statistics, a mixture distribution is the probability distribution of a random variable whose values can be interpreted as being derived in a simple way from an underlying set of other random variables. In particular, the final outcome value is selected at random from among the underlying values, with a certain probability of selection being associated with each. The underlying random variables may be random vectors, each having the same dimension, in which case the mixture distribution is a multivariate distribution. In other words, the mixture model expresses the probability distribution function of a random variable as the sum of probability densities of other random variables. A general linear combination of probability density functions is not necessarily a probability density, since it may be negative or it may integrate to something other than 1. However, a convex combination of probability density functions preserves both of these properties (non-negativity and integrating to 1), and thus mixture densities are themselves probability density functions. Hence, the description of a

mixture of distributions is straightforward; any convex combination

$$g(x) = \sum_{i=1}^{n} \alpha_i f_i(x), \qquad \sum_{i=1}^{n} \alpha_i = 1 \tag{2.33}$$

of other probability distribution $f_i(x)$, is a "mixture distribution model" [143].

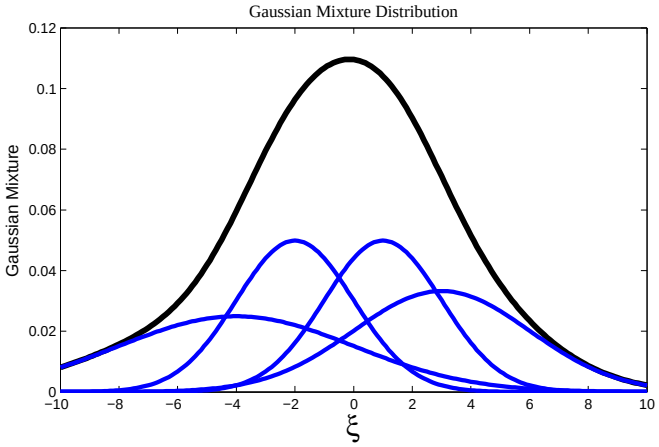

Figure 2.6: A Gaussian mixture model

The particular case is when the $f_i(x)$'s are all normal distributions $\mathcal{N}(\mu, \sigma^2)$, and $g(x)$ is written as the sum of Gaussian distribution. In statistics, the mixture of Gaussian distributions has been widely used for density estimation and data smoothing problem where inferences about the population are made based on a finite data sample [166]. Figure 2.6 shows a mixture model of Gaussian distributions.

A distinction needs to be made between a random variable whose probability distribution function is the sum of a set of components, e.g. a mixture distribution, and a random variable whose value is approximated by the sum of functions of other random variables, e.g. polynomial chaos expansion, in which case the distribution is given by the convolution operator. For instance, the sum of two normally-distributed random variables, each with different means, will still be a normal distribution. On the other hand, a mixture density created as a mixture of two normal distributions with different means will have two peaks provided that the two means are far enough apart, showing that this distribution is radically different from a normal distribution.

Gaussian mixture density estimates are closely related to normalized histograms, but endowed with properties such as smoothness or continuity. To illustrate this, assume we are given a data set whose histogram corresponds to Figure 2.7. We use the following Gaussian mixture model to estimate the probability distribution of given data.

$$g(x) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-2)^2}{2}} + \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+2)^2}{2}}. \tag{2.34}$$

Accordingly we obtain an approximation for the probability density function of the

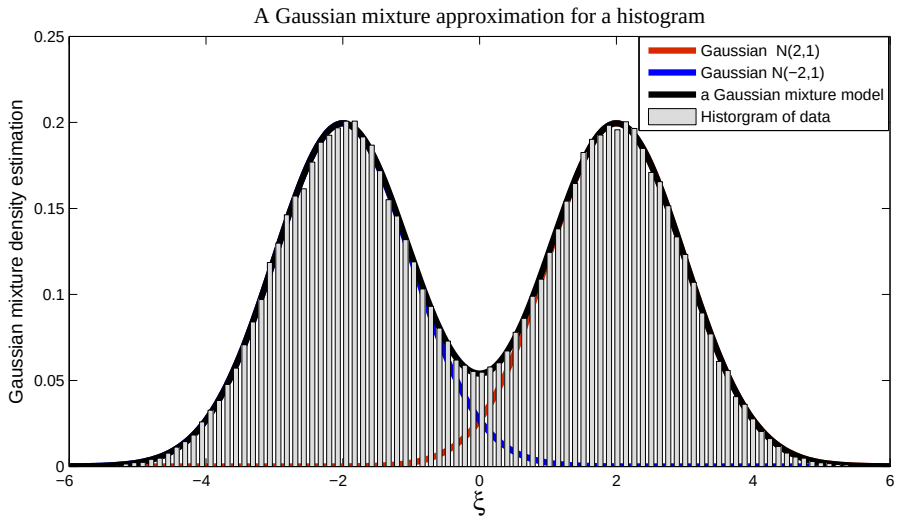

Figure 2.7: Gaussian mixture density estimate of a histogram

aforementioned histogram by using a Gaussian mixture model. At this stage we have an expression for the distribution of the underlying random variable, and in the next step we employ the standard Hermite polynomial chaos representation of $g(x)$ to obtain a random variable whose value is approximated by sum of the Hermite polynomials of a Gaussian random variable, $\xi \sim \mathcal{N}(0,1)$. Although $g(x)$ is created by the mixture of two normal distributions with different means, x is not a Gaussian distribution. Following the discussion of Section 2.3, its probability space needs to be transformed to the same probability space of the Gaussian random variable ξ. By using the equation (2.20), we have:

$$u = \int_{-\infty}^{X} g(x)dx = \int_{-\infty}^{\xi} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi. \tag{2.35}$$

Assuming $u \in [0,1]$ has a uniform distribution, by computing the inverse cumulative distribution of x and ξ, their probability space can be transformed to the probability space of u;

$$X = G^{-1}(u) \qquad and \qquad \xi = W^{-1}(u). \tag{2.36}$$

Figure 2.8 illustrates the procedure for the probability space transformation. Having both ξ and x in the same probability space, it is now straightforward to calculate the coefficients of the polynomial chaos expansion, by computing the integral of the equation (2.19) numerically. Table 2.2 shows the values for the polynomial chaos coefficients of

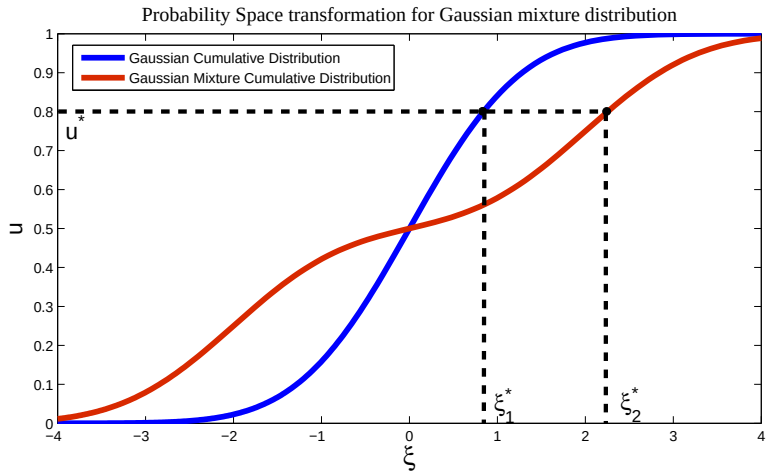

Figure 2.8: The Gaussian mixture probability space transformation of Example 2.3.2

different orders.

$$PC_n(\xi) = \sum_{i=1}^{n} a_i H_i(\xi) = -0.001 + 2.1813(\xi) - 0.0422(\xi^2 - 1) - 0.1766(\xi^3 - \xi) + \ldots. \tag{2.37}$$

Figure 2.9 compares the probability distribution of the polynomial chaos expansions of different orders with the Gaussian mixture distribution of the equation (5.2).

Table 2.2: The polynomial chaos coefficients for the Gaussian mixture distribution

a_0	-0.0010
a_1	2.1813
a_2	-0.0422
a_3	-0.1766
a_4	-0.0139
a_5	0.0277
a_6	-0.0004

Figure 2.9 demonstrates the convergence in distribution of the polynomial chaos approximation to the Gaussian mixture distribution of the equation (5.2), as the order of expansion increases. However, the rate of convergence is slow comparing to Example 2.3.1, where the underlying random variable was Gaussian. In fact, convergence of any functional in L_2 does not necessarily mean a fast convergence of the polynomial chaos expansion for the functionals of non-Gaussian random variables. When the underlying random variables are Gaussian, polynomial chaos expansion achieves a fast convergence rate of the exponential order $\mathcal{O}(\frac{1}{(p+1)!})$, as stated by Cameron-Martin [25] in Section 2.3. But when it comes to the functionals of non-Gaussian random variables,

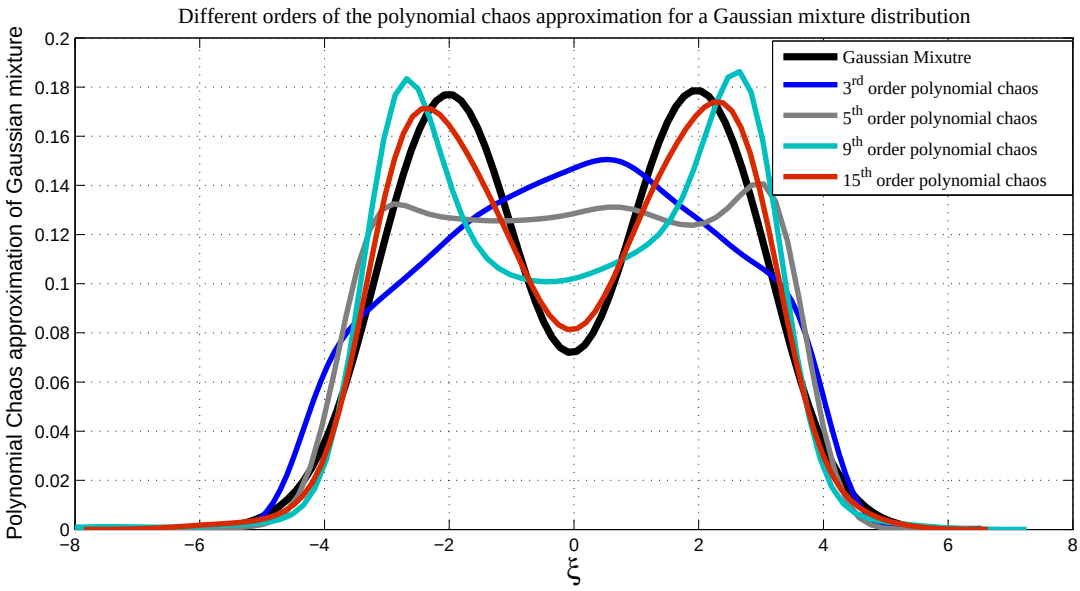

Figure 2.9: Different orders of the polynomial chaos approximation of the Gaussian mixture probability distribution function of the equation (5.2)

e.g. Gaussian mixtures, the convergence rate of the standard Wiener polynomial chaos expansion might be moderate. To resolve this issue, the generalized polynomial chaos was introduced by Xiu and Karniadakis [199] to enhance the convergence rate of the standard Wiener polynomial chaos expansion for functionals of non-Gaussian random variables, which will be discussed in more detail in Section 2.5.

2.4 Numerical Methods to Compute Polynomial Chaos Coefficients

As mentioned earlier, the greatest benefit of the polynomial chaos expansion as apposed to any other polynomial or other approximations is that it guarantees convergence in probability as the order of the polynomial chaos expansion increases. This implies that it does not have the problem of over-fitting that plagues many other proxies. However, there is a caveat, that is, for the theoretical convergence results to be valid, the coefficients of polynomial chaos expansion have to be estimated correctly. Several numerical methods have been proposed in the literature to compute the coefficients accurately. They can be classified into two categories of *intrusive methods* and *non-intrusive methods.*

Intrusive methods estimate the coefficients in a way that the residual is orthogonal

to the polynomial chaos basis and require access to the equations of the mathematical model to compute the polynomial chaos coefficients (i.e., a "black box" approach is not possible). On the other hand, Non-intrusive methods use a solution set (trial runs) to approximate the polynomial chaos coefficients. The main benefit of non-intrusive methods is they allow for a "black box" approach. For a given number of trial runs, the coefficients associated with higher order polynomial chaos terms have larger estimation errors compared to lower order terms, and therefore, there is a balance between the number of trial runs and the maximum possible order of polynomial that can be reliably used. In the following, we briefly discuss intrusive and non-Intrusive methods.

2.4.1 Intrusive Methods

Intrusive methods find the polynomial chaos coefficients using the governing equation of the model. Galerkin projection scheme is the most common algorithm for the intrusive calculation of the polynomial chaos coefficients.

Galerkin Projection Scheme

Stochastic Galerkin method transforms a stochastic equation into a set of coupled deterministic equations [64]. Consider the general stochastic partial differential equation

$$\mathcal{L}(x,t;\theta) = f(x,t;\theta), \tag{2.38}$$

where $\mathcal{L}(x,t;\theta)$ is a linear or nonlinear operator, $f(x,t;\theta)$ is the source term and the random process $u^*(x,t;\theta)$ is the analytical solution. The random parameter θ represents uncertainty introduced by initial or boundary conditions, rate parameters, system properties, etc. Using the polynomial chaos expansion we can approximate $u^*(x,t;\theta)$ by $\tilde{u}(x,t;\theta)$ as

$$u^*(x,t;\theta) \approx \tilde{u}(x,t;\theta) = \sum_{i=0}^{P} u_i(x,t)\mathbf{H_i}(\xi(\theta)). \tag{2.39}$$

Here we have restricted the infinite sum of the equation (2.9) to a finite summation involving $P+1$ expansion terms. The total number of terms depends on the number of dimensions (N) of the multivariate random parameter ξ and also the highest order (M) of the orthogonal polynomials set $\{\mathbf{H_i}\}$ according to the required accuracy [64]

$$(P+1) = \frac{(N+M)!}{(N!)(M!)}. \tag{2.40}$$

By substituting the truncated spectral expansion of u^* into the model equation (2.38), we can compute the residual as

$$\hat{e} = \mathcal{L}(\tilde{u}, t; \theta) - f(\tilde{u}, t; \theta), \tag{2.41}$$

where $\tilde{u}(x, t; \theta) \approx \sum_{i=0}^{P} u_i(x, t)\mathbf{H_i}(\xi(\theta))$. The Galerkin projection scheme determines the polynomial chaos coefficients in a way that the residual of equation (2.41) is orthogonal to the polynomial chaos basis $\{\mathbf{H_i}\}$;

$$\begin{aligned}
&\hat{e} \perp \mathbf{H_k} \qquad \forall k = 0, \ldots, P \\
&\Rightarrow \mathbf{E}[(\mathcal{L}(\tilde{u}, t; \theta) - f(\tilde{u}, t; \theta))(\mathbf{H_k}(\xi(\theta)))] = 0 \\
&\Rightarrow \mathbf{E}[(\mathcal{L}(\sum_{i=0}^{P} u_i(x, t)\mathbf{H_i}(\xi(\theta)), t; \theta) - f(\sum_{i=0}^{P} u_i(x, t)\mathbf{H_i}(\xi(\theta)), t; \theta))(\mathbf{H_k}(\xi(\theta)))] = 0.
\end{aligned} \tag{2.42}$$

The Galerkin projection scheme guarantees that the error in the approximate solution is orthogonal to the space spanned by $\{\mathbf{H_i}\}$. By using the orthogonality property of Hermite polynomial basis $\{\mathbf{H_i}\}$, the stochastic differential equation of (2.38) reduces to a $P + 1$ system of coupled deterministic differential equations. We can use any appropriate spatial and temporal discretization of the coefficients to numerically solve this system and compute the polynomial chaos coefficients $u_i(x, t)$.

The main disadvantage of the Galerkin method is that it requires the access to the governing equations of the system and does not allow for the black-box approach.

2.4.2 Non-Intrusive Methods

Non-intrusive methods use a solution set (trial runs) to approximate the polynomial chaos coefficients. The accuracy of the estimation depends on the choice of selection of the solution set and also the number of trial runs. Different choices of the trial runs results in different accuracy of the estimation of the polynomial chaos coefficients. The non-intrusive approach treats the system as a black box. The conventional choice for the trial runs (collocation points) is the random selection. However, an accurate and consistent solution generated by the randomly chosen points is not guaranteed for each run. Moreover, the alignment or clustering of the data points sometimes causes the insufficient information to estimate the polynomial chaos coefficients. Hence, the trial runs needs to be sampled intelligently from places that will provide more information about the stochastic process under study. Based on the sampling strategy

for collocation points and the numerical method to compute the polynomial chaos coefficients consequently, Non-intrusive methods can be categorized into the *probabilistic collocation method* and the *spectral projection method*. In this section, we briefly review these methods.

Probabilistic Collocation Method

The general collocation method belongs to a family of mathematical techniques that reduces a complex model to a simpler model, for example they transform differential equations to algebraic equations [177]. The basic concept of the probabilistic collocation method is to try to approximate the response of the model $u^*(x,t;\theta)$. Our presentation of the collocation method in the following section closely follows the exposition of Sarma [152].

From the equation (2.39) we write the truncated polynomial chaos expansion for the response model,

$$\tilde{u}(x,t;\theta) \approx \sum_{i=0}^{P} u_i(x,t)\mathbf{H_i}(\xi(\theta)). \tag{2.43}$$

Here, $\{\mathbf{H_i}\}$ is the set of polynomial chaos orthogonal basis in terms of the independent random variables ξ, and $\tilde{u}$ is the approximation of u^*. By evaluating the true forward model $P+1$ times, the equation (2.43) can be used to derive a set of linear equations that can be solved to determine $u_i(x,t)$. The quality of approximation depends preponderantly on the choice of values of input parameters at which the forward model is evaluated. These values are called "the collocation points". The collocation points have to be selected in a way that the equation (2.43) captures as much information about u^* as possible. In most problems, prior information about the uncertain input parameters is available. Hence, the collocation points must be selected in such a way that they cover the high probability regions of these probability distributions. In order to achieve this, one idea is to use the Gaussian quadrature technique [192] for estimating integrals. The Gaussian quadrature approach uses the roots of the next higher order polynomial to estimate the integral of a polynomial [192].

To illustrate the idea of the probabilistic collocation method, consider the probability density function of random variable A shown in Figure (2.10). To estimate a linear approximation of $y = f(A)$, heuristically we require two points that span the high probability region of $P(A)$, as illustrated in figure (2.10). These two points are basically the roots of the second order orthogonal polynomials that span $P(A)$. By using the two points bounding the high probability region of $P(A)$, an accurate estimate $\tilde{y}$ for $y = f(A)$ can be obtained. The larger deviation of $\tilde{y}$ from y only occurs in the low probability

region and it thus contributes only a small error. The orthogonal polynomials can be

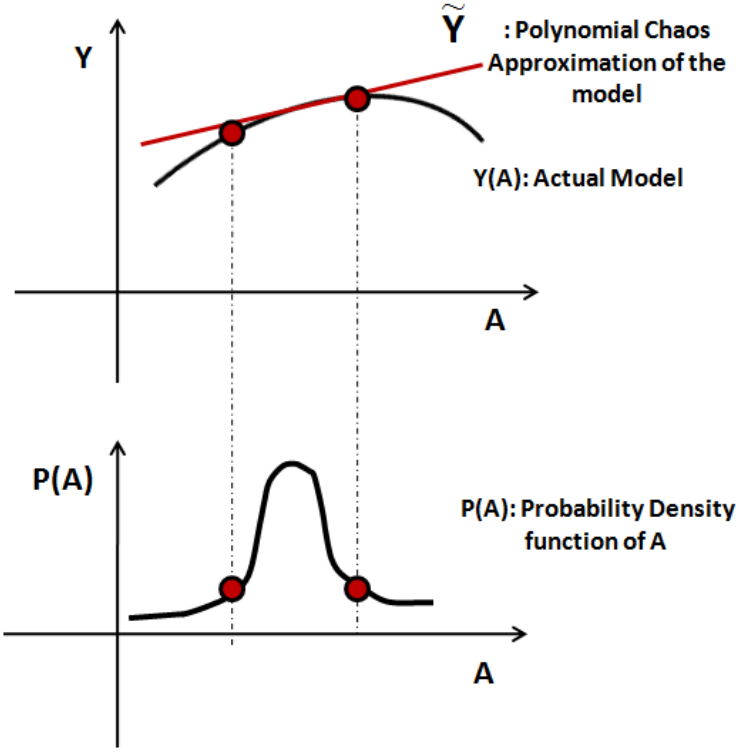

Figure 2.10: For a linear approximation of a function, the two collocation points are chosen in a way that span a high probability region of the probability distribution of random input.

constructed numerically with the respect to the probability distribution of input random variable. Further, the number of terms in the polynomial expansion of the equation (2.43) should be as small as possible (while providing a good approximation of the forward model at the same time), because the number of true model evaluations is equal to $P + 1$.

In general, the number of roots is usually much larger than the number of collocation points required, especially when the number of input random variables is large. For example, the roots of the third order Hermite polynomial, $\{0, \sqrt{3}, -\sqrt{3}\}$, provide the collocation points for the model runs of the second order polynomial chaos expansions. Accordingly, for a two dimensional problem, the possible choices for collocation points are $(0,0), (\sqrt{3},\sqrt{3}), (-\sqrt{3},-\sqrt{3}), (\sqrt{3},-\sqrt{3}), (-\sqrt{3},\sqrt{3}), (0,\sqrt{3}), (0,-\sqrt{3}), (\sqrt{3},0), (-\sqrt{3},0)$. There are nine possible choices for collocation points but from the equation (2.40) with $N = 2$ and taking terms only up to $\mathbf{H}_2$ ($M = 2$), there are only six unknowns. Similarly, for high dimensional problems, the number of available collocation choices is much larger than the required number of collocation points, which poses a problem of

selection the optimal collocation points. Figure 2.11 illustrates the collocation points for the Hermite polynomial of order 10, with two random variables ($N = 2, M = 10$) using the Gaussian quadrature technique. The selection criteria is usually based on the probability of the points in a way that the more probable choices are selected first. This method is regarded as an optimal method for one-dimensional problems, because it gives the same results as the Galerkin method [76]. Once the $P+1$ collocation points are chosen, the forward model is evaluated $P+1$ times. Using the equation (2.43) and the $P+1$ true responses, $P+1$ linear equations can be derived, which can be solved to obtain the $P+1$ coefficients $u_i(x,t)$. More details on different methods of the collocation points selection, e.g. Sobol sequence [123], and the optimality discussion can be found at [76].

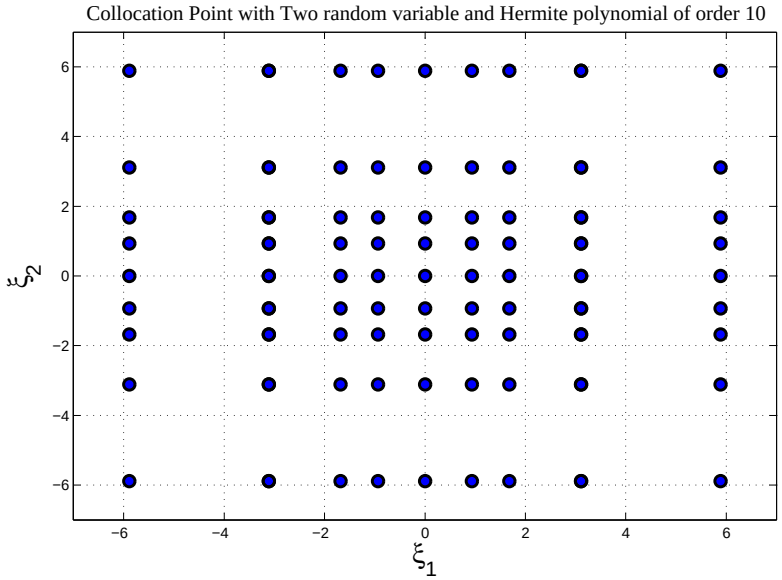

Figure 2.11: The collocation points for the Hermite polynomial $\mathbf{H_{10}}(\xi_1, \xi_2)$, with two random variables using the Gaussian quadrature technique.

For a reliable use of the polynomial chaos expansion, the convergence quality of the truncated expansion has to be examined first. For this purpose, we run the forward model a few more times and compare the true model results with the polynomial chaos approximation results. We select these points for the forward model run from the next higher order polynomial chaos expansion. The next higher order is used because if the errors are too large and a higher order of approximation is required, we will already have the model solutions needed to solve the approximation. Hence, if the error is small enough, the approximation can be used for uncertainty analysis, otherwise, the next higher order approximation is evaluated and the process repeated until convergence is

achieved.

There is one caveat with the probabilistic collocation method that the input random variables have to be independent. In the case when the random input properties consist of random fields (a set of spatially distributed correlated random variables) or random processes (a set of temporally distributed correlated random variables), the standard probabilistic collocation method cannot be used directly due to the independence assumption of the input random variables ξ [76]. This however is usually the case with reservoir simulation models, whose input parameters such as porosity and permeability of one grid cell are related to those of another grid cell. The Karhunen-Loeve expansion, as it will be explained in the next chapter, gives an approximation for a random field in terms of a set of independent random variables while maintaining the covariance structure. Thus, once the Karhunen-Loeve expansion of the input random field is obtained, the probabilistic collocation method can then be performed using the independent random variables resulted from the Karhunen-Loeve expansion. Further, the number of transformed independent random variables is much smaller than the initial correlated random variables, which is very advantageous for the probabilistic collocation method as the number of collocation points and therefore the efficiency depends directly on the number of input variables. Refer to Chapter 3 for a description of the Karhunen-Loeve expansion.

However, there are some key limitations in the application of the standard collocation method to large scale simulation models. In the traditional probabilistic collocation method, the number of simulations required is equal to the number of coefficient (terms) in the polynomial chaos expansion, which grows exponentially with the number of random variables and order of the expansion. As mentioned earlier, in the standard collocation method, the polynomial chaos coefficients are obtained by solving a linear system of equations of size $P + 1$, where $P + 1$ is the number of coefficients, which is given by the equation (2.40). We see that even for a relatively small problem with 10 variables and 5^{th} order polynomials, the number of terms is 3003, which is clearly impractical to run as many simulations for realistic problems.

Isukapalli [76] proposed the **regression-based PCM** as a modified probabilistic collocation method where he applies a least-square technique to determine the polynomial chaos coefficients, and instead of solving a linear system a regression problem is solved. As such, the number of trial runs need not be equal to the number of polynomial chaos coefficients in regression-based PCM. However, if the number of simulations is too small compared to the number of terms, the regression problem is severely undetermined and non-unique. For a given number of trial runs, in the regression-based PCM,

the coefficients associated with higher order polynomial chaos terms have larger estimation errors compared to lower order terms, and therefore, there is a balance between the number of trial runs and the maximum possible order of polynomial that can be reliably used.

To mitigate the "curse of dimensionality", one approach is to only use "pure" terms of the polynomial chaos expansion and discard the cross terms [100, 156]. This reduces the number of coefficients drastically and seemed to give good results on some examples tested in [100]. However, this will work only if the true forward model does not contain such nonlinear cross-effects, which is not generally true for reservoir simulation models [156]. Even more generally, using this approach invalidates one of the biggest benefits of polynomial chaos expansion, which is the guaranteed convergence to the true distribution as the order of the polynomial chaos expansion increases.

Non-Intrusive Spectral Projection

The basic idea behind the non-intrusive spectral projection method to find the polynomial chaos coefficients ($u_i(x,t)$), is to use the orthogonality property of the polynomial chaos basis, i.e. $E[\mathbf{H_i}(\xi(\theta))\mathbf{H_j}(\xi(\theta))] = 0;\ \forall i \neq j$ [142, 156]. By applying the orthogonality condition to the truncated polynomial chaos approximation of a general stochastic process $u^*(x,t;\theta)$, we obtain:

$$\begin{aligned} u^*(x,t;\theta) &\approx \sum_{i=0}^{P} u_i(x,t)\mathbf{H_i}(\xi(\theta)) \\ E[u^*(x,t;\theta)\mathbf{H_i}(\xi(\theta))] &\approx u_i(x,t)E[\mathbf{H_i^2}(\xi(\theta))] \\ \Rightarrow u_i(x,t) &= \frac{E[u^*(x,t;\theta)\mathbf{H_i}(\xi(\theta))]}{E[\mathbf{H_i^2}(\xi(\theta))]}. \end{aligned} \tag{2.44}$$

The denominator in the above equation is an expectation of known polynomial chaos basis and can therefore be calculated analytically. The expectation in the numerator involves $u^*(x,t;\theta)$ and therefore, the estimation of $u_i(x,t)$ basically amounts to calculation of this expectation accurately. Calculating this expectation can be posed as a sampling problem or an N dimensional numerical integration problem, and in both cases, it can be written as:

$$E[u^*(x,t;\theta)\mathbf{H_i}(\xi(\theta))] \approx \frac{1}{N_s}\sum_{k=1}^{N_s} w_k u^*(x,t;\theta_k)\mathbf{H_i}(\xi(\theta_\mathbf{k})). \tag{2.45}$$

Here N_s is the number of samples (number of simulations), and w_k are weights for each sample point, which depends on the method used. For example, if standard Monte

carlo sampling is used, then $w_k = 1$ and the samples are obtained randomly, whereas, if the numerical integration techniques such as quadrature is used, these weights are different and the samples are distributed uniformly in the integration domain. It is clear that N_s is not directly dependent on the number of terms in the polynomial chaos expansion $P+1$, and we only need as many simulations as necessary to reduce the error in the estimate of $u_i(x,t)$ to an acceptable level. The second benefit of the non-intrusive spectral projection method is that once the N_s simulations are evaluated, each coefficient $u_i(x,t)$ is evaluated independently of other coefficients, as opposed to the probabilistic collocation method where all the coefficients are evaluated by solving a coupled linear system or a regression problem. This implies that in the spectral projection method, the estimate of one coefficient is not dependent on that of another coefficient, and this can be very important. If they are dependent as in probabilistic collocation method, then the estimate of higher order coefficients (usually less accurate) will affect the estimate of the lower order coefficients, thereby also affecting their accuracy [156].

In order to determine the best sampling approach in order to estimate $u_i(x,t)$ in the equation (2.44), the rate of convergence of the estimation error for the expectation $E[u^*(x,t;\theta)\mathbf{H_i}(\xi(\theta))]$ with N_s samples has to be studied. For usual numerical integration techniques such as quadrature and cubature, unfortunately, the error converges at the rate $\mathcal{O}(\frac{1}{N_s^{\frac{c}{d}}})$, where d is the number of random variables and c is a positive constant dependent on the particular algorithm. Since the rate of convergence is a function of d, it is not practical to use these methods when $d > 4$ or 5 [156]. On the other hand, the standard Monte Carlo method has the convergence rate of $\mathcal{O}(\frac{1}{\sqrt{N_s}})$, which is independent of the dimension d. This however, is a slow rate of convergence requiring typically thousands of reservoir simulation runs, which may not be practical [156].

A more efficient approach is to use *Qausi Monte Carlo* using low discrepancy sequence [122, 156], which been shown to have a convergence rate of $\mathcal{O}(\frac{1}{N_s})$ for large class of problems [14]. For a given number of simulations N_s, the error in the estimate of $u_k(x,t)$ depends on the order of polynomial chaos term $\mathbf{H_k}$ associated with it. The higher the order of $\mathbf{H_k}$ is, the larger error will be in the estimate of $u_k(x,t)$ for a given N_s. If a coefficient has significant error, then keeping it in the truncated polynomial chaos expansion will reduce the accuracy of the expansion. However, the attraction of the approach is that, because this coefficient is estimated using sampling method with N_s simulation runs, rather than using a data fitting approach as in the probabilistic collocation method, the same simulations can be used in a technique similar to the cross-validation method to determine whether the estimated coefficient is sufficiently accurate to retain it, or should it be discarded. Sarma et. al [156] studied

the application of Quasi Monte Carlo spectral projection method with Sobol sequence [122], for uncertainty quantification in reservoir simulation and demonstrated the linear convergence rate when the cross-validation technique is used.

2.5 Generalized Polynomial Chaos Expansions

The classical Wiener-Hermite polynomial chaos expansions are based on the Hermite polynomial functionals in terms of Gaussian random variables and they have an exponential convergence rate. In theory, they converge to any second order functional on the random space [25]. However, in practice the convergence rate of the Wiener polynomial chaos expansion for non-Gaussian random field is slow [199]. Moreover, it can not be applied to random field with discrete probability distribution. Accordingly, to improve the convergence rate of the polynomial chaos expansion for non-Gaussian random variables, the coordinates in probability space, where the basis of the polynomial chaos are constructed, must be adapted to the statistics of the input random field and the random solution. Xiu and Karniadakis [199] proposed the Askey-chaos as a generalization of the original Wiener polynomial chaos expansion. In the Askey chaos, the underlying random variables are not limited to Gaussian random variables and the orthogonal polynomial basis of the Askey chaos are selected adaptively from the Askey scheme of orthogonal polynomials, with the respect to input random variable

Similar to the Section (2.3) we represent a general second-order random process $X(\theta)$ as

$$\begin{aligned} X(\theta) = c_0 I_0 + \sum_{i_1=1}^{\infty} c_{i_1} I_1(\xi_{i_1}(\theta)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} c_{i_1 i_2} I_2(\xi_{i_1}(\theta), \xi_{i_2}(\theta)) \\ + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} c_{i_1 i_2 i_3} I_3(\xi_{i_1}(\theta), \xi_{i_2}(\theta), \xi_{i_3}(\theta)) + \dots, \end{aligned} \tag{2.46}$$

where $I_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$ denotes the Askey chaos of order n in terms of the multi-dimensional random variables $\xi = (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$. Comparing the above representation to the equation (2.9), demonstrates that in the Askey chaos, the orthogonal polynomial basis I_n are not restricted to Hermite polynomials and could be any type of orthogonal polynomials from the Askey scheme, as illustrated in Table 2.3.

The orthogonality condition for the polynomial basis of the Askey chaos implies that:

$$\langle I_i I_j \rangle = \langle I_i^2 \rangle \delta_{ij}, \tag{2.47}$$

where the ensemble average denotes the inner product of the equation (2.13) in the Hilbert space of the random variable ξ under the measure $w(\xi)$. Here, the weighting function $w(\xi)$ corresponds to the Askey chaos basis I_i. The weighting function for some types of the Askey orthogonal polynomials are the same as the probability distribution of certain types of random variables. Accordingly, these random variables are chosen as the independent underlying random variables ξ in the equation (2.46).

Table 2.3: The correspondence between the type of the generalized polynomial chaos and their underlying random variables ($N \geq 0$ is a finite integer) [199].

	Random Variables ξ	Askey Chaos $\mathbf{\Phi_i}(\xi)$	Support
Continuous	Gaussian	Hermite-Chaos	$(-\infty, \infty)$
	Gamma	Laguerre-Chaos	$[0, \infty)$
	Beta	Jacobi-Chaos	$[a, b]$
	Uniform	Legendre-Chaos	$[a, b]$
Discrete	Poisson	Charlier-Chaos	$\{0, 1, 2, \ldots\}$
	Binomial	Krawtchouk-Chaos	$\{0, 1, \ldots, N\}$
	Negative Binomial	Meixner-Chaos	$\{0, 1, 2, \ldots\}$
	Hypergeometric	Hahn-Chaos	$\{0, 1, \ldots, N\}$

Table 2.3 illustrates that when the underlying random variable is Gaussian, the Askey chaos recovers the standard Wiener polynomial chaos expansion. It also classifies the Askey chaos expansion into continues Askey chaos and discrete Askey chaos that represents random variables with discrete probability distributions. The Legendre polynomial is actually a special case of the Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ where $\alpha = \beta = 0$, and it corresponds to the uniform distribution. The Legendre chaos is listed separately in Table 2.3, because of the importance of the uniform distribution.

The convergence properties of the Askey-chaos is studied in [199], and exponential convergence rate was computationally demonstrated computationally for model problems. The aforementioned correspondence can be extended to arbitrary probability distributions with the orthogonal polynomials constructed on-the-fly; this extension was presented in [187]. They proposed a multi-element generalized polynomial chaos method to deal with stochastic inputs using an arbitrary probability measures. Based on the decomposition of the random space of the stochastic inputs, a set of orthogonal polynomials is numerically constructed with respect to a conditional probability density function in each element and subsequently the generalized polynomial chaos is implemented locally [188].

2.6 Summary

We have reviewed the constructions of the standard Wiener polynomial chaos expansions of random variables with finite second moments, and illustrated that under the condition of the Cameron and Martin theorem, the polynomial chaos expansion guarantees the convergence in probability as the order of the expansion increases. We presented some convergence properties of general real-valued random variables, then utilized these properties to make observations regarding polynomial chaos approximations for the L_2 random variables.

The greatest benefit of polynomial chaos expansion as opposed to any other polynomial or other approximations is that it guarantees convergence in probability as the order of the polynomial chaos expansion increases. This implies that it does not have the problem of over-fitting that plagues many other proxies. However, there is a caveat, that is, for the theoretical convergence results to be valid, the coefficients of polynomial chaos expansion have to be estimated correctly. We described the Intrusive and Non-Intrusive methods as a general classification of numerical methods to estimate the polynomial chaos coefficients. The main advantage of non-intrusive methods over intrusive methods is the feasibility for the black-box approach. On the other hand, the main disadvantage of the non-intrusive methods is the high number of function evaluations for the higher order of polynomial chaos expansion.

Galerkin methods presented as a class of intrusive methods that require access to the equations of the mathematical model to compute the polynomial chaos coefficients (i.e., a black-box approach is not possible). It estimates the coefficients in a way that the residual becomes orthogonal to the polynomial chaos basis. Among non-intrusive methods, the probabilistic collocation method and the spectral projection technique were also studied. They use a solution set (trial runs) to approximate the polynomial chaos coefficients. For a given number of trial runs, the coefficients associated with higher order polynomial chaos terms have larger estimation errors compared to lower order terms, and therefore, there is a trade-off between the number of trial runs and the maximum possible order of the polynomial chaos that can be reliably used.

To achieve faster convergence rate of the polynomial chaos expansion for non-Gaussian random variables, the Askey chaos is introduced as a generalized polynomial chaos expansion. In the Askey chaos expansion, the polynomial basis are chosen from the Askey scheme of orthogonal polynomials in accord with the underlying random variables. For any given type of random input, the Askey chaos converges in probability but the exponential rate of convergence is only retained if the probability distribution

of random input is similar to the weighting function of the Askey chaos expansion.

Chapter 3

Efficient Polynomial Chaos Proxy-based History Matching

3.1 Introduction

There exits a potential for large improvements in oil recovery by using a more accurate reservoir model. Developing a plan to maximize oil production requires constructing reservoir models constrained to all available data. Reservoir modeling is, however, still a vexed question because of various sources and types of data that need to be integrated as well as the possibly existing uncertainty due to lack of data to fully constrain the reservoir model. From today's oil fields, many types of data are being obtained. One of the most important data is provided by geologists. Geologists produce a geological interpretation of the reservoir from outcrop or other inspections, resulting in e.g predictions of channel dimensions, their stacking patterns or where the turbulent flow in the ocean is located. Additionally, direct observation from a few wells is available as a form of well log, core, or well test data. On the other hand, indirect observations from geophysical surveys (especially seismic survey), often termed "soft" data, provide lower-resolution constraints. Additionally, production history (bottom hole pressure, oil or water rate) is recorded during the production. Matching the reservoir model to the production history is generally very difficult due to the severe nonlinearity between the reservoir model and the history. History matching problems are generally ill-posed and may have non-unique solutions [22]. To reduce the ill-posedness of the problem and have a reservoir model with a reasonable predictive quality, it is necessary to integrate all available data during the process of creating a history matched model.

In practice, reservoir model parameters, such as permeability are only known with a large uncertainty and consequently even history-matched reservoir models are merely a rough approximation of reality. In particular at the appraisal stage, where reservoir production data is few and where critical decisions need to be made, uncertainty about reservoir volume and prediction performance is still considerable and critical to the decision making process. Therefore the predictive value of such reservoir models is limited. To improve the predictive quality of the reservoir model, it has to be perpetually updated during its life time to assimilate into hitherto measured production data.

History matching techniques have been studied widely in the petroleum engineering context for the last three decades, and various algorithms have been developed to perform history matching. Chapter 1 presents a brief review of the existing history matching techniques. A comprehensive literature review of the history matching techniques in the reservoir engineering context can be found at Oliver et al. [126].

In practice, most history matching analysis is still performed manually or at most uses assisted history matching techniques [118]. This may be due to the fact that the existing history matching techniques are either hard to implement or they suffer from number of drawbacks, such as not being able to honor geological constraints [20]. Hence new developments are still required to acquire a reliable and robust method for automatic closed-loop history matching.

Existing history matching algorithms can be roughly classified into two general categories: stochastic algorithms and gradient-based methods. Gradient-based algorithms have several inherent limitations, including the need to compute the gradients at each step of the optimization process and also the possibility of converging to a local minimum (rather than global minimum). Moreover, most of the existing gradient-based history matching algorithms, do not take geological constraints into account [153]. An important advantage of stochastic algorithms is that they are conveniently capable of honoring complex geological constraints by preserving multi-point statistics present in the prior geological model. They are also very easy to implement and converge to the global minimum [197, 103]. However, the main disadvantage of the stochastic methods is their inefficiency, as they require large number of simulations for convergence [153]. Therefore, it can be extremely time-consuming for high resolution reservoir models. This is particularly of concern in closed-loop reservoir management [78, 155], which requires continuous real-time use of history matching algorithms. Thus, there is a significant need for an efficient proxy (or surrogate) model that can predict simulation results with reasonable accuracy.

Dimension reduction techniques, which have been applied in many application areas

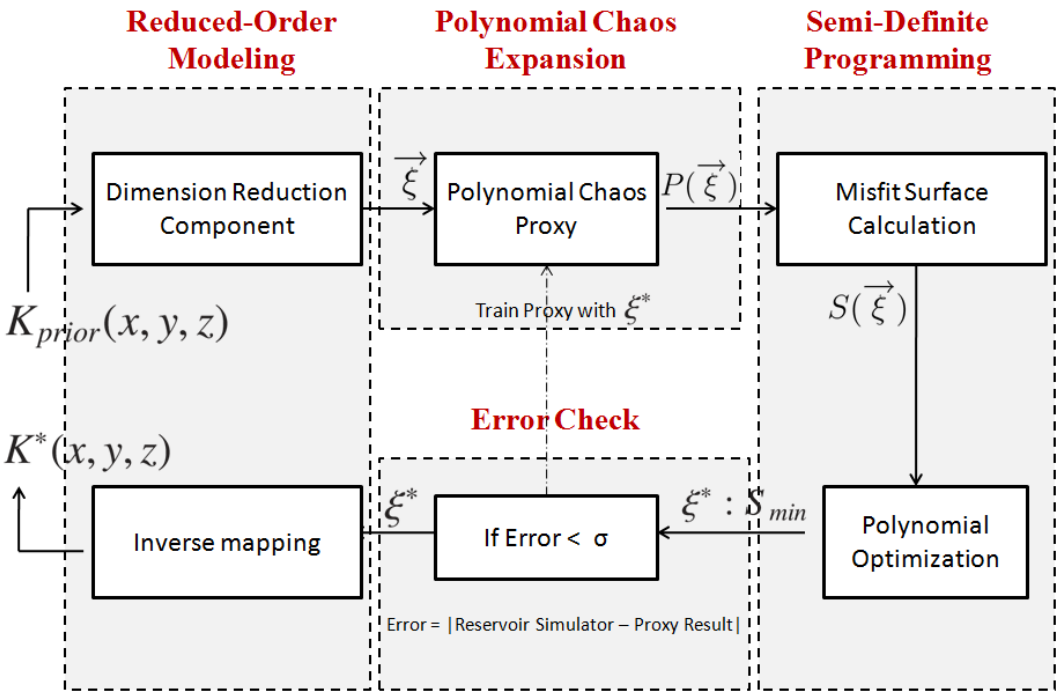

Figure 3.1: Polynomial chaos proxy-based history matching framework

including reservoir simulation, represent a promising means for constructing efficient surrogate models. Many of these techniques entail the projection of the high resolution description of the reservoir into a low-dimensional subspace, which significantly reduces the number of unknowns. The use of the Karhunen-Loeve representation was first introduced by Gavalas et al. [59] within a reservoir engineering context to reduce the dimension of geological parameters. The basic approach was later used by Oliver [125], Reynolds et al. [145], and Sarma et al. [155]. This enables us to approximate the high resolution geological model with the lower dimension of space. Although Karhunen-Loeve is very useful and effective in preserving the correlation structure of the geological model, for preserving higher-moments of non-Gaussian random fields, Kernel Principal Component Analysis (KPCA) was used by Sarma et al. [154]. It was shown in their work that complex geological structures can be effectively decomposed into lower dimension by mapping the random field into feature space.

Polynomial chaos expansion, which was discussed in Chapter 2, constitutes a promising means to construct an efficient pseudo-simulator (proxy). Polynomial chaos expansion has a significant advantage over other response surfaces and proxy models as it guarantees the convergence in probability and also in distribution to the output random variable of interest, i.e. cumulative oil production. Convergence of polynomial

chaos expansion in probability to the full reservoir simulator results implies that it can be accurately yet efficiently used as a proxy for the reservoir simulator. The use of polynomial chaos expansion for uncertainty quantification was pioneered by Ghanem and Spanos [64] and has been applied to various engineering problems. Sarma et al. [155] used polynomial chaos expansion to efficiently quantify uncertainty for closed-loop production optimization and model updating. Another significant advantage of using the polynomial chaos expansion as a proxy is that despite other stochastic methods of optimization, the global minimum of multi-variate polynomials can be very efficiently computed via current polynomial optimization techniques [19]. Semi-definite programming (SDP) which is a relatively new field of optimization with growing interest, is a subfield of convex optimization concerned with the optimization of a well-behaved functions over the intersection of the cone of positive semi-definite matrices. Subsequently, SDP can be efficiently applied to optimize multi-variate polynomials which can be written in terms of sum of squares of polynomials. Figure 3.1 illustrates the proposed framework for automatic history matching procedure using the polynomial chaos proxy. It entails three main blocks, the component for dimension reduction of the geological models, the component of polynomial chaos expansion and the component of semi-definite programming for minimization of the misfit function. In the following, we will elaborate each component in detail.

This chapter is structured as follows; first we review a summary of the geological parameterization methods for a reservoir model. Then we explain the dimension reduction component in Figure 3.1 and study Karhunen-Loeve decomposition and kernel PCA as two efficient tools for dimension reduction. Accordingly, we denote the use of polynomial chaos expansion as a pseudo-simulator where it is trained with the reduced-order parameters of the geological model. History matching is then posed as a minimization problem of the misfit surface and we will briefly review the recently developed techniques of semi-definite programming for polynomial optimization to efficiently find the global minimum. In the remainder of the chapter, history matching problem of a synthetic fluvial-channel case has been studied using the proposed polynomial chaos proxy.

3.2 Geological Parameterization

For reservoir modeling, data is very difficult and usually expensive to get and may require indirect measurements of physical reservoir properties. Data sparsity is a function of the cost involved in collecting subsurface data by measurement of the reservoir.

Measurements are typically taken directly from within the well or indirectly via a geophysical measurement technique. Wells are the only route to measuring the actual reservoir rocks. Well costs vary greatly depending on whether they are located onshore or offshore, the technical difficulty of the well and rig and personnel costs, but are typically in the millions of pounds range. As such only a small number of wells can be drilled and consequently only a small number of measurements can be taken. A number of measurement devices are available to source data from the wellbore, and either measure the reservoir unit properties and types (i.e. petrophysical methods, corelogs), reservoir fluid properties (i.e. repeat formation testers (RFT), PVT samples) or reservoir dynamic responses and pressures (i.e well test analysis, downhole pressure gauges). Not all tests will be performed on each well and some may not be performed at all due to the costs involved.

Geologists deal with the sparsity of data by using prior knowledge about what is and is not geologically possible to reduce the number of possible models. These expert judgements are based on the experience of the geologist in inferring probabilities about the unknown data using different but related data sources. An example of this may be to infer porosity and net/gross values to estimate hydrocarbon volumes for an undrilled exploration well, based on previously drilled wells in the region or outcrops of reservoir facies exposed at the surface.

The fundamental steps to develop a reservoir models entails successively building:

- the reservoir geometry
- the geological model
- the petrophysical model.

The reservoir geometry model is usually identified by subdividing the field into homogenous units corresponding to different depositional environments. Each unit shares the same statistical properties. Outcrops can be used to distinguish different layers and bedding. The main constraints results from the intercepts of the wells with these geological units. The seismic knowledge can also help to identify the lithofacies units and faulting information. Figure 3.3 illustrated different depo-

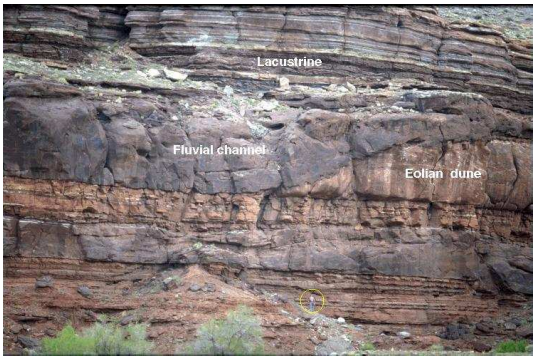

Figure 3.2: Outcrop used for geological interpretation

sitional layers constrained by well intercepts. In geostatistical modeling, the lithofacies units and the petrophysical quantities are all defined as random variables. Each random variable is characterized by its spatial behavior.

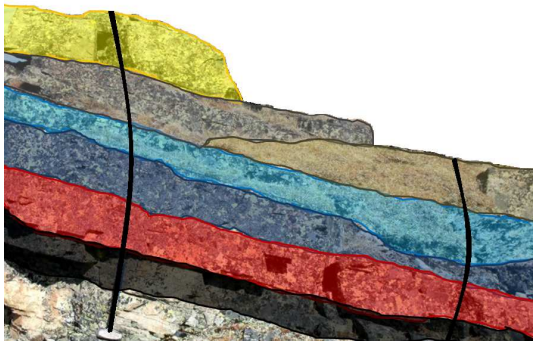

Figure 3.3: layers constrained by well intercepts

In geostatistics, the term *spatial relationship* describes many forms of relations among the available data and various unknowns [23]. The data may be of any type, possibly different from that of the variable being estimated. Therefore, to apply geostatistics, one needs to first and for most quantify that spatial relationship through a geological continuity model. The decision of which data to pool together to provide the estimate for statistical properties of a random variable requires, in geostatistical term, a *decision of stationarity*. It would not make sense to estimate the porosity at a specific location on the basis of data that originates from a different layer or fault block in the reservoir with completely different geological (and, hence, porosity) characteristics. In statistical terms, the data must have similar statistical properties. However, one should avoid the temptation to make every single datum value unique! Some form of pooling data is necessary for statistical methods to work.

The underlying geological continuity makes those data values that are "closer" to the target location more informative than data farther away; hence close data should get more attention in the geostatistical model. The question here lies in the definition of "closeness". Simply defining distance as the Euclidean distance between two points would ignore the specificity of geological continuity. To illustrate this, consider Figure 3.4 at which the statistical properties of point Z is estimated using data values at D_1, D_2, D_3. Although the datum at D_1 is farther from the unknown location Z than that at D_2 and D_3, the underlying diagonal continuity makes the value at D_1 more informative. One important contribution of geostatistics is that it incorporates information about spatial continuity into statistical estimation.

On the other hand, data values that are sampled close to each other are somewhat redundant; hence, each value of such a cluster is less valuable than an isolated datum at the same distance from the unknown. Geostatistics takes this principle into account by putting more weight on the data that carries more information [23].

The traditional geostatistical methods use a two-point statistics information known

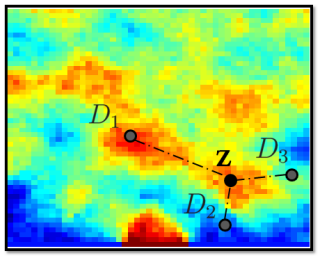

Figure 3.4: The underlying geological continuity determines that the data at location D_1 is more relevant to the estimation of the unknown at Z than the datum at locations D_2 or D_3

as *variogram*. Variogram is a simple but effective way to investigate the spatial variation of the variable of interest. It is defined as the variance of the difference between random field values at two locations. The use of variogram as a rectilinear estimation of parameters makes the result of geostatistical simulation relatively smooth [23]. Since anisotropy and curvilinearity are common characteristics of geological structures, the variogram measure of connectivity can not preserve the certain distribution pattern of complex geological features such as fractures, channels, fluvial, delta, etc. A suggestion for development was multi-point geostatistics, in which for example, a training image, a 2D matrix template is used for inference of spatial variation. It was shown in the literature that training images are useful to dictate the desired patterns while honoring geological constraints [23]. Section 3.3 will review multi-point geostatistics in more detail.

Many random variables might be involved in reservoir modeling. However, some parameters are usually superior to the others in the sense that they have a dominant impact on fluid flow. *Parameterization* can be simply defined as the description of a complex system in terms of a discrete set of physically interpretable quantities that capture the key elements of that system. The process of developing a model parameterization is to identify the most important model components in terms of impact on flow and degree of uncertainty, then define prior ranges for each of these parameters. Several geological parameterization methods have been developed recently by a number of authors to increase the geological validity of history matching. Silva et al. [36] applied a number of techniques to automated history matching of different reservoirs from around the world. In each case the key reservoir uncertainties were identified and a suitable parameterization was developed to reflect realistic variations in the reservoir geology. In the case of Silva et al. the parameterization were an attempt to capture geological uncertainty but predominantly dealt with intrinsically non-geological param-

eters such as permeability multipliers between layers. Key to defining an appropriate parameterization is an adequate description of the parameter prior distributions and knowledge of the likely probability distributions of those parameters, which may or may not conform to regular distribution shapes such as the uniform or Gaussian. This requires an understanding of the sensitivity of reservoir parameters, knowledge of which parameters are the least certain, and the combination of all available prior knowledge that can used to describe the probability of the uncertain parameters.

For the remainder of this chapter, without loss of generality, we consider geological parameterizations in which only permeability field varies from model to model but porosity is constant. Hence, reservoir model updating involves tuning permeability field in a way that gives the minimum misfit value with the observed production data of the reservoir under study.

3.3 Dimensionality Reduction

When two or more random variables relate to each other, the correlation between them enables us to reduce the dimension. As for geological models, the rock and fluid properties in different locations usually exhibit some correlation to each other and therefore one can efficiently reduce the number of parameters without much geological information. The curse of dimensionality makes the reservoir proxy models less efficient, hence it is essential to reduce the number of unknown geological parameters for the history matching problem. Figure 3.5 demonstrates how correlation between the permeability of two adjacent grid cells of a simulation can bring about the feasibility of reducing the unknown parameters. The more correlated a random field is, the more effectively it can be represented in a reduced dimension space. Even complex geological structures that exhibit correlation or connectivity between their geological parameters can be effectively decomposed into lower dimensions.

Traditional geostatistics uses a variogram, $2\gamma(x, y)$, to measure geological continuity and it is defined as the variance of the difference between random field values at two locations [23]:

$$2\gamma(x, y) = \mathrm{var}(Z(x) - Z(y)) = E\left(|(Z(x) - \mu(x)) - (Z(y) - \mu(y))|^2\right). \tag{3.1}$$

If the lithofacies unit under study has a constant mean μ, the equation (3.1) takes the form:

$$2\gamma(x, y) = E\left(|Z(x) - Z(y)|^2\right), \tag{3.2}$$

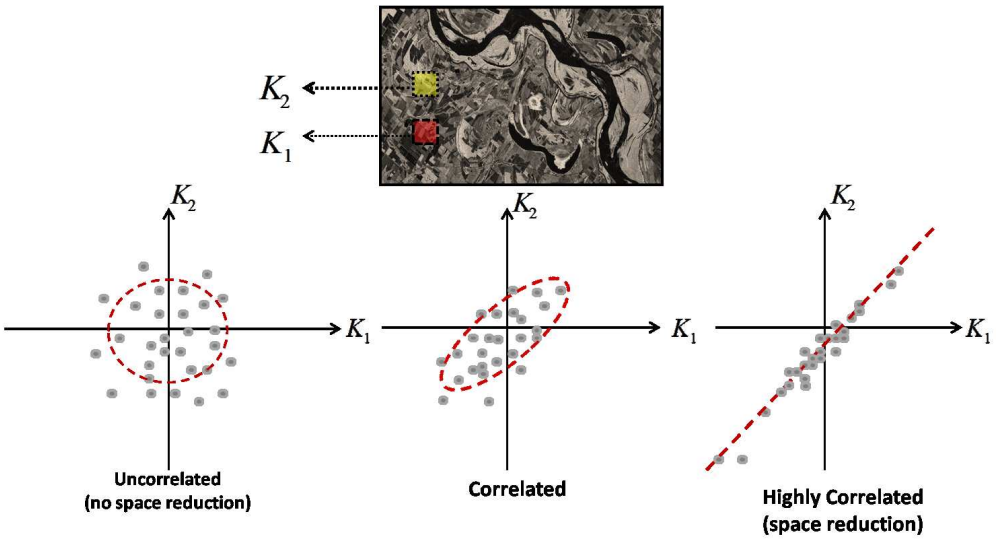

Figure 3.5: Various degrees of correlation between K_1 and K_2

where $\gamma(x, y)$ itself is called the *semivariogram*. If a random field is stationary in a wide sense, the variogram can be expressed as a function of the difference $h = |y - x|$ between locations as:

$$\gamma_s(y - x) = \gamma_s(h). \tag{3.3}$$

In this case for observations $D_i,\ i = 1, \ldots, k$ at locations $x_1, \ldots, x_k$ the empirical variogram $\hat{\gamma}(h)$ can be calculated as [23]:

$$\hat{\gamma}(h) := \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} |D_i - D_j|^2, \tag{3.4}$$

where $N(h)$ illustrates the set of observations pairs $i,\ j$ where $|x_i - x_j| = h$. In practice, the empirical variogram can be used as an approximation for the theoretical variogram in the presence of enough geological information. To ensure validity of the empirical variograms in applied geostatistics, they are usually approximated by model functions [23].

Because the variogram can only capture the correlation of two points located in distance h (two-point statistics), it cannot model curvilinear structures such as channels, nor can it model strong patterns of conductivities such as fractures. The representation of such complex geological features requires multi-point statistics, involving jointly more than two locations. The idea behind multi-point geostatistics is to infer spatial patterns using many spatial locations [23]. Figure 3.6 illustrates the difference between two-point based inference and multi-point geostatistics.

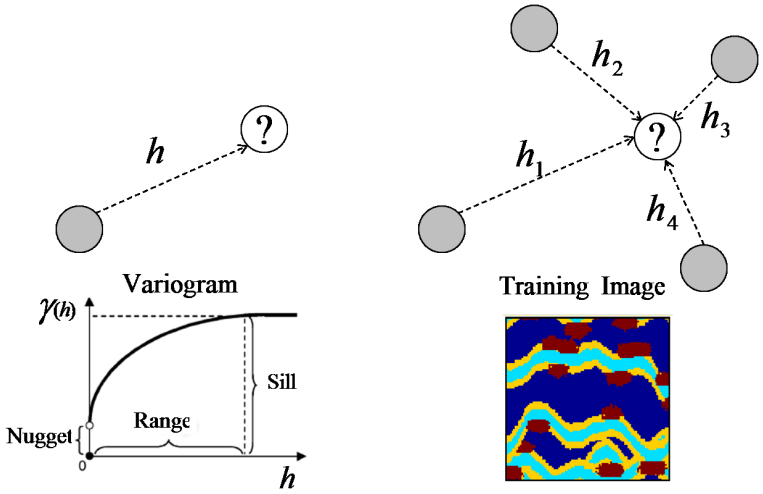

Figure 3.6: The two points configuration used to infer statistics in traditional variogram-based approaches (left), while a sample of a multi-point (5-point) configuration used in multiple-point geostatistics to model highly complex geological phenomenon and curvilinear sedimentary features using heuristic training images (right).

The training image is a heuristic reservoir analogue that imitates the geological heterogeneity and contiguous patterns of conductivity in the lithofacies under study [86, 173]. Training images are merely conceptual images to capture high-order statistics and they are not required to be constrained to any subsurface data. In applied geostatistics, Boolean algorithm are commonly used to generate training images [86, 173].

Strebelle [173] proposed *snesim* algorithm to generate realizations constrained to reservoir data. The snesim algorithm is basically similar to the traditional Gaussian sequential simulation that generates the value of one grid cell after another, but the probability model is different. In traditional method, the probability model from which the value of a grid cell is generated, is assumed to be a Gaussian distribution with mean and variance determined by a set of variogram based *kriging* equations [23], while in the snesim algorithm, the probability model is inferred from the training image rather than a variogram model. Refer to Strebelle [173] for more detail on sequentially generating geostatistical models based on the training images.

Preserving multi-point statistics using a training image enables us to account for nonlinear spatial relationships and generate realizations drawn from the prior joint probability distribution of geological parameters. This can be better illustrated if we consider Figure 3.7 where two geological parameters, permeability at two different grid cells, is plotted against each other for different realizations. For reservoir models with a rectilinear spatial relationship, where the variogram captures the latent dependencies of

parameters, the plot exhibits a straight line. However, for complex geological structures where a training image based algorithm is used to generate realizations, the plot shows a nonlinear dependency. It is worth noting that by "complex" geological structure, we mean it consists of random geological parameters that have highly nonlinear correlations, not uncorrelated geological parameters. There is no room for dimensionality reduction where the geological parameters are totally uncorrelated.

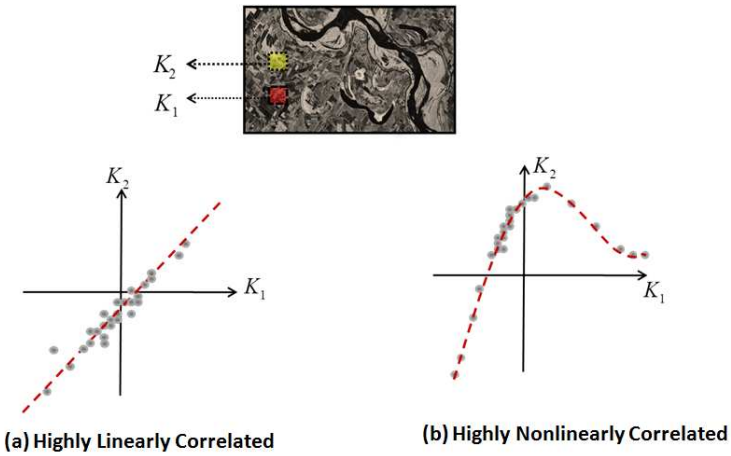

Figure 3.7: Complex geological structures such as channels generally have highly nonlinear correlation

A key issue with both the traditional two-point and the more recent multi-point geostatistical algorithms is that the algorithms do not provide a differentiable functional relationship between the output geological realizations and the input parameters of the algorithms. Hence we have to take grid cells, geological properties as random parameters. In practice, reservoir models usually have 10^6-10^9 grid cells and it is impossible to build a reservoir surrogate model on these high-dimensional geological parameters. Reduced-order modeling mitigates the curse of dimensionality and renders a low-dimensional representation of geological parameters. Various techniques for linear and nonlinear reduced-order modeling have been developed in the literature.

In the petroleum engineering community, Gavalas et al. [59] used the Karhunen-Loeve (K-L) expansion or linear principal component analysis (PCA) for a differentiable linear reduced-order modeling of subsurface properties in terms of a small number of independent random variables. This representation was also used later by Oliver [125] and Reynolds et al. [145] for the history-matching problem. However, the K-L expansion can only retrieve covariance of the random field (two-point statistics), and therefore is suitable only for multi-Gaussian random fields. It cannot be applied to model complex geological structures such as channels. For non-Gaussian random fields, kernel PCA was

used by Sarma et al. [154] to reduce dimensions of complex geological structures. Kernel PCA is a nonlinear generalization of the K-L expansion that preserves the multi-point statistics of non-Gaussian random fields. In the following we explain the K-L expansion and kernel PCA as powerful tools to express high-dimensional geological parameters in terms of a small number of independent random variables.

Reduced-order modeling only focuses on non-supervised methods which attempts to eliminate any redundancy in the initial variables and give a low-dimensional representation of geological parameters. The supervised methods of dimensionality reduction was discussed in Section 3.2 where geologists describe a complex geological structure in terms of a discrete set of physically interpretable quantities that capture the key elements of that system. Accordingly, reduced-order modeling aims to just detect and eliminate the dependencies between the given geological parameters.

3.3.1 The Karhunen-Loeve Expansion of Random Fields

The Karhunen-Loeve expansion is an effective tool for representing random field (stationary or non-stationary) with known covariance function, in terms of a set of uncorrelated random variables and deterministic functions [64]. The deterministic functions are basically the eigenfunctions of covariance function that decay steadily. Hence, it provides a second moment characterization of a random field in terms of uncorrelated random variables and orthogonal deterministic functions. One of the important features of the K-L expansion is the bi-orthogonality property of both random variables and eigenfunctions that allows for an optimal encapsulation of information contained in multi-Gaussian random process into a set of independent random variables [75].

If the analytical covariance function is not readily available, we can use the ensemble covariance matrix to approximately perform the Karhunen-Loeve transformation. Given a set of realizations of a random field $y_k,\ k = 1, \ldots, N_R,\ (y_k \in R^{N_C})$ the ensemble covariance matrix can be calculated as [161]

$$\mathbf{C} = \frac{1}{N_R} \sum_{j=1}^{N_R} (y_j - \bar{y})(y_j - \bar{y})^T, \tag{3.5}$$

where N_R is the number of realizations and N_C is the dimension of the random field. The number of realizations N_R should be large enough to yield a converged covariance matrix $\mathbf{C}$. If the variance of the random field is large, we may need tens of thousands of realizations to reach to a converged covariance matrix. This approach of calculating the ensemble covariance matrix numerically from a set of realizations of the random field as above, instead of using some analytical covariance model, is the most general approach

for calculating the covariance. For any arbitrary random field, an analytical covariance model may not exist, but the numerical approach can always be applied. For example, if the number of realizations are created using a multi-point geostatistical algorithm such as *snesim* [173], the covariance matrix can always be calculated numerically from those realizations; however, it is possible that an analytical covariance model associated with those realizations may not exist.

The discrete form of the K-L expansion with the covariance $\mathbf{C}$ is given as [75]:

$$y = \bar{y} + \mathbf{E}\Lambda^{\frac{1}{2}}\xi, \tag{3.6}$$

where $\mathbf{E}$ is the matrix of eigenvectors of the covariance matrix $\mathbf{C}$, Λ is a diagonal matrix of the eigenvalues and ξ is a set of uncorrelated random variables. If the random field is multi-Gaussian, the underlying random variables will be independent Gaussian random variables. Hence, the K-L expansion can be effectively used for representing multi-Gaussian random fields. In contrast, for non-Gaussian random fields, the equation (3.6) can still be used to approximate the realizations y_k, but the higher order moments will not be reproduced and the approximated realizations only have the same covariance as the original realizations.

In fact, the K-L expansion is a parameterization of the form $y = f(\xi)$, where the functional relationship is linear. Here, matrix $\mathbf{C}$ is of the size $N_C \times N_C$. The maximum size of the matrices $\mathbf{E}$ and Λ is $N_C \times N_C$, and that of vector ξ is $N_C \times 1$. The word maximum is used because we may choose to retain only the largest N_M of the total N_C eigenvalues, in which case $\mathbf{E}$ is of size $N_C \times N_M$, Λ is of size $N_M \times N_M$, and ξ is of size $N_M \times 1$. Furthermore, the maximum number of non-zero eigenvalues is actually the minimum of N_C and N_R, implying that N_C non-zero eigenvalues do not exist if $N_R < N_C$ [154]. Finally, the high-dimensional random field y is approximately represented in terms of low-dimensional N_M independent random variables ξ. There is a map from the high-dimensional space R^{N_C} to the low-dimensional space R^{N_M}. Figure 3.8 illustrates schematically the dimensionality reduction concept by K-L expansion.

Discarding the $(N_C - N_M)$ smallest eigenvalues implies that we are discarding the shortest correlation lengths (higher frequency changes). This will lead to the loss of some fine-scale geological information. However, the K-L expansion is an optimal expansion for multi-Gaussian random fields in a least-square sense [75]. This means that of all possible dimensionality reduction methods of the multi-Gaussian random field with N_M random variables, the K-L expansion minimizes the least-square approximation error. Even when the random field is not multi-Gaussian, the K-L expansion minimizes the least-square approximation error of representing the realizations used to

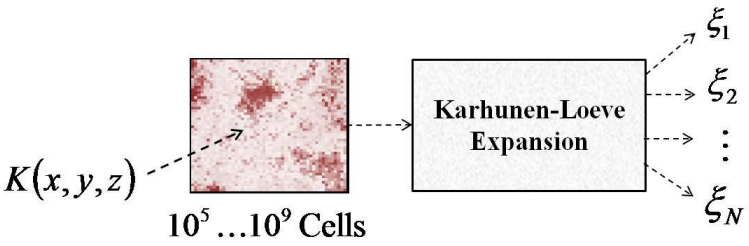

Figure 3.8: Reduced-order representation by the Karhunen-Loeve expansion

create the covariance matrix $\mathbf{C}$. This is a favorable property of the K-L expansion for the applications discussed earlier, because a relatively small number of random variables N_M is essential for efficiency of proxy models, though this still provides accurate approximations of multi-Gaussian random fields.

In order to determine the K-L expansion, the following eigenvalue problem has to be solved [161, 154]:

$$\lambda \mathbf{v} = \mathbf{C}\mathbf{v}, \tag{3.7}$$

where λ is an eigenvalue and $\mathbf{v}$ is the corresponding eigenvector of the covariance matrix $\mathbf{C}$. The computational cost of solving the eigenvalue problem of the equation (3.7) using standard methods (such as singular value decomposition method) is a very expensive process of $\mathcal{O}(N_C^3)$ complexity [66]. Hence, it is infeasible to solve the eigenvalue problem for large-scale reservoir simulation models. Sarma et al. [154] proposed an alternative but equivalent formulation of the same problem, called the kernel eigenvalue problem, which can be solved much more efficiently to solve for the non-zero eigenvalues λ and eigenvectors $\mathbf{v}$ of the ensemble covariance matrix $\mathbf{C}$. However, for a covariance model with large variances, it is cumbersome to apply the kernel eigenvalue method as tens of thousands of realizations are required to reach a converged ensemble covariance matrix.

To solve the eigenvalue problem for an analytically given covariance model with large variance, we propose a random projection method which uses the *Johnson-Lindenstrauss lemma* [84] to directly compute the eigenvalues of eigenvectors very efficiently. In the following we first explain the kernel eigenvalue problem for ensemble covariance matrix and then the Johnson-Lindenstrauss lemma of random projection to find the eigenvalues and eigenvectors of an analytical covariance model.

- **Kernel eigenvalue problem of an ensemble covariance matrix**
  Given a set of realizations of a random field $y_k,\ k = 1, \ldots, N_R,\ (y_k \in R^{N_C})$, we use the definition of the ensemble covariance matrix of the equation (3.5) and the eigenvalue problem of the equation (3.7), to obtain the following formulation

[161, 154],

$$\mathbf{Cv} = \frac{1}{N_R}\sum_{j=1}^{N_R}(\widetilde{y_j}.\mathbf{v})(\widetilde{y_j}), \tag{3.8}$$

where $\widetilde{y_j} = (y_j - \bar{y})$ is a vector of size $N_R \times 1$,$\mathbf{v}$ is the corresponding eigenvector of the covariance matrix $\mathbf{C}$. This implies that all solutions $\mathbf{v}$ with $\lambda \neq 0$ must lie in the span of the N_R realizations of $y_1, y_2, \ldots, y_{N_R}$. Since eigenvectors $\mathbf{v}$ must lie in the span of N_R realizations, and there cannot be more than N_R orthogonal directions in the span of N_R realizations, there can only be N_R non-zero eigenvalues associated with these N_R eigenvectors. Therefore, $\mathbf{C}$ can have a maximum of only N_R non-zero eigenvalues if $N_R < N_C$. This has two important consequences [161, 154]. First, the equation (3.7) can be written in the following equivalent form

$$\lambda(y_k.\mathbf{v}) = (\widetilde{y_k}.\mathbf{Cv}) \qquad \forall\, k = 1, \ldots, N_R, \tag{3.9}$$

and second, there exist coefficients α_j such that

$$\mathbf{v} = \sum_{j=1}^{N_R}\alpha_j\widetilde{y_j}. \tag{3.10}$$

By combining the equation (3.9) and (3.10), we obtain [154]:

$$\lambda\sum_{i=1}^{N_R}\alpha_i(\widetilde{y_k}.\widetilde{y_i}) = \frac{1}{N_R}\sum_{i=1}^{N_R}\alpha_i(\widetilde{y_k}.\sum_{j=1}^{N_R}\widetilde{y_j})(\widetilde{y_i}.\widetilde{y_i}) \qquad \forall\, k = 1, \ldots, N_R. \tag{3.11}$$

Defining an $N_R \times N_R$ matrix $\mathbf{K}$ where $\mathbf{K_{ij}} = (\widetilde{y_i}.\widetilde{y_j})$ is the dot product of two vectors corresponding to realizations i and j, the equation (3.11) can be written as [154]:

$$N_R\lambda\mathbf{K}\alpha = \mathbf{K^2}\alpha. \tag{3.12}$$

$\mathbf{K}$ is called the kernel matrix, and it is of the size $N_R \times N_R$, while the covariance matrix $\mathbf{C}$ is of size $N_C \times N_C$. Here $\mathbf{K_{ij}} = (\widetilde{y_i}.\widetilde{y_j})$ is called the polynomial kernel of order 1 [161, 154]. As described in Scholkopf et al. [161], it can be shown that the eigenvalues and eigenvectors of the eigenvalue problem of the equation (3.12) are equivalent to those of the following eigenvalue problem

$$N_R\lambda\alpha = \mathbf{K}\alpha. \tag{3.13}$$

The equation (3.13) is known as the kernel eigenvalue problem. The eigenvalues are given by $N_R\lambda$ and the eigenvectors are given by α. Solving this problem is exactly equivalent to solving the equation (3.7), because the non-zero eigenvalues of the equation (3.12) are just that of (3.7) scaled by N_R, and the eigenvectors $\mathbf{v}$ associated with the non-zero eigenvalues λ of (3.7) can be obtained from α by using (3.12). The rest of the $(N_C - N_R)$ eigenvalues of $\mathbf{C}$ are equal to zero [154]. The attractiveness of solving the kernel eigenvalue problem, instead of the original problem of the equation (3.7), is that for practical problems $N_R \ll N_C$ and consequently the eigenvalue problem can be solved extremely efficient. The only limitation of this method is that the number of realizations N_R should be large enough to yield a converged ensemble covariance matrix $\mathbf{C}$. For covariance models with high variances, one needs to generate tens of thousands of high-dimensional realizations to achieve a converged ensemble covariance matrix.

- **Johnson-Lindenstrauss method**

  The Johnson-Lindenstrauss (JL) lemma [84] asserts that an n-point set in any Euclidean space can be mapped to a Euclidean space of dimension $k = \epsilon^{-2}\log(n)$ in a way that all distances are preserved up to a multiplicative distortion factor between $[1-\epsilon]$ and $[1+\epsilon]$ [39].

  The JL lemma is somewhat surprising as it states the feasibility to compress the information about the distance between n-points in space, into a very much smaller matrix. But the construction of the map is perhaps even more surprising. Known proofs obtain such a mapping as a super-simple linear map $\mathbf{W} : R^n \rightarrow R^k$, where $\mathbf{W}$ is an orthonormal random matrix [84]. It was shown such a map preserves all distances with probability $P \geq [1 - \frac{1}{n^2}]$. When n is very large, $\mathbf{W}$ preserves all distances almost surely with the small distortion ϵ.

  As an immediate consequence of the JL lemma, If we define the distance between N_C random variables as the covariance in L^2 space, $d = \mathbf{E}(|x-y|^2)$, the two-point statistics can be preserved with small distortion ϵ by random projection into a much lower dimensional space $N_M = O(\epsilon^{-2}\log(N_C))$. Since the K-L expansion only captures the two-point statistics of a random field, the projection into sub-space will approximately keep the principal eigenvalues the same. To elaborate JL lemma, let $\mathbf{X} : (x_1, \ldots, x_{N_C})$ be a centered random field in multidimensional space and $\mathbf{W} : R^{N_C} \rightarrow R^{N_M}$ be any orthonormal random matrix, the linear projection of $\mathbf{X}$ into low-dimensional space will be $\mathbf{Y} = \mathbf{W^T X}$. The covariance matrix of the random field $\mathbf{Y}$ can be computed as:

$$\mathbf{C_Y} = \mathbf{E}[\mathbf{Y}\mathbf{Y}^\mathbf{T}] = \mathbf{E}[\mathbf{W}^\mathbf{T}\mathbf{X}\mathbf{X}^\mathbf{T}\mathbf{W}] = \mathbf{W}^\mathbf{T}\mathbf{E}[\mathbf{X}\mathbf{X}^\mathbf{T}]\mathbf{W} = \mathbf{W}^\mathbf{T}\mathbf{C_X}\mathbf{W}. \tag{3.14}$$

The dimension of $\mathbf{C_Y}$ is $(N_M \times N_M)$ where $N_M \ll N_C$. Fowler [57] proved that the principal eigenvalues of $\mathbf{C_Y}$ are a reasonable approximate of the principal eigenvalues of $\mathbf{C_X}$ with the error bound stated by the Rayleigh-Ritz theorem [134], and the principal eigenvectors of Υ_X and Υ_Y are related to each other by

$$\Upsilon_X = \mathbf{W}\Upsilon_Y. \tag{3.15}$$

Hence, we can solve the eigenvalue problem of equation (3.7) in a much lower dimensional space, and compute the principal eigenvalues and eigenvectors of high-dimensional space by equation (3.15).

Example 3.3.1. *We study a stationary Gaussian random field* $\mathbf{Y}$ *whose number of random variables (cells) is* $N_C = 3600$*. The mean of random field is zero* $(\mathbf{E}[\mathbf{Y}] = 0)$ *and the covariance matrix of the random field is given by the Gaussian covariance model:*

$$\mathbf{C}(\overrightarrow{a}, \overrightarrow{b}) = \sigma^2 e^{-\frac{(a_x - b_x)^2}{\lambda_x} - \frac{(a_y - b_y)^2}{\lambda_y}} \tag{3.16}$$

where λ_x *and* λ_y *are the dimensionless correlation lengths in the direction of* x *and* y*, and* σ^2 *is the variance of the stationary random field. To write the Karhunen-Loeve expansion for the random field under study, we need to solve the eigenvalue problem for the covariance function of equation (3.16). As mentioned earlier, solving the eigenvalue problem here is an expensive process of* $\mathcal{O}(N_C^3)$ *complexity. To efficiently find the eigenvalues and eigenvectors for the analytically given covariance function (the Gaussian model), we use JL lemma for random projection to compute the largest eigenvalues. Let* W *be an orthogonal random matrix of size* (3600×100)*, the random projection of the exponential covariance matrix to a lower-dimensional can be computed via the equation (3.17):*

$$\mathbf{C_R} = \mathbf{W}^\mathbf{T}\mathbf{C}\mathbf{W}. \tag{3.17}$$

The size of $\mathbf{C_R}$ *is* (100×100) *and the eigenvalue problem can be efficiently solved to find its eigenvalues and eigenvectors. The 100 largest eigenvalues of* $\mathbf{C}$ *is approximately equal to the eigenvalues of* $\mathbf{C_R}$*. The eigenvectors corresponding to these eigenvalues can be computed by the equation (3.15). Figure 3.9 shows the 100 largest eigenvalues of the Gaussian covariance function, where* $\sigma^2 = 0.5$ *and* $\lambda_x = \lambda_y = 0.25$*.*

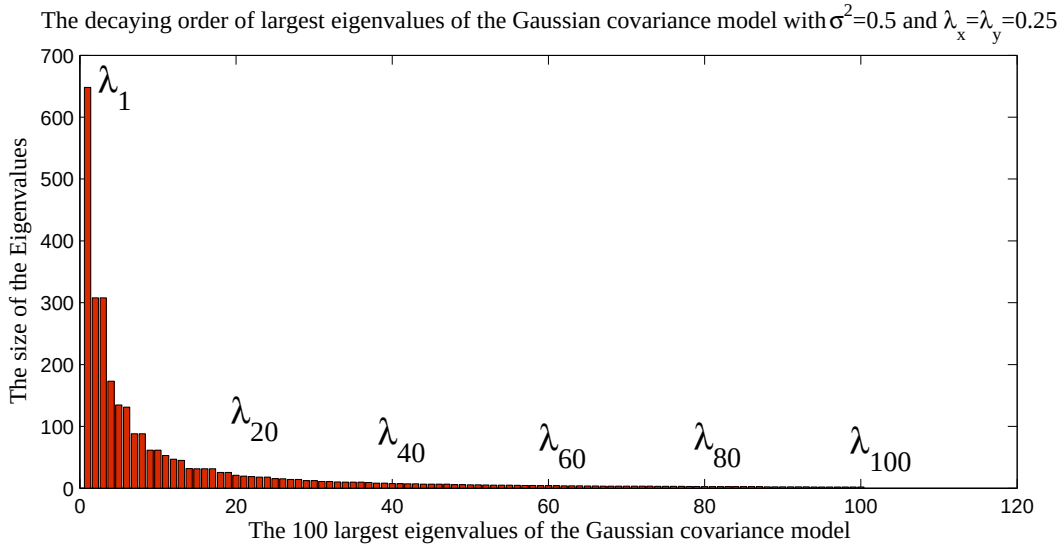

Figure 3.9: The largest eigenvalues of the Gaussian covariance function, $\sigma^2 = 0.5$ and $\lambda_x = \lambda_y = 0.25$

Figure 3.10 illustrates the eigenfunctions corresponding to the 10 largest eigenvalues in Figure 3.9. The higher-order eigenfunctions capture the higher frequency changes in the Karhunen-Loeve representation of the random field. This can be better illustrated in Figure 3.11 where the higher-order eigenfunctions are depicted.

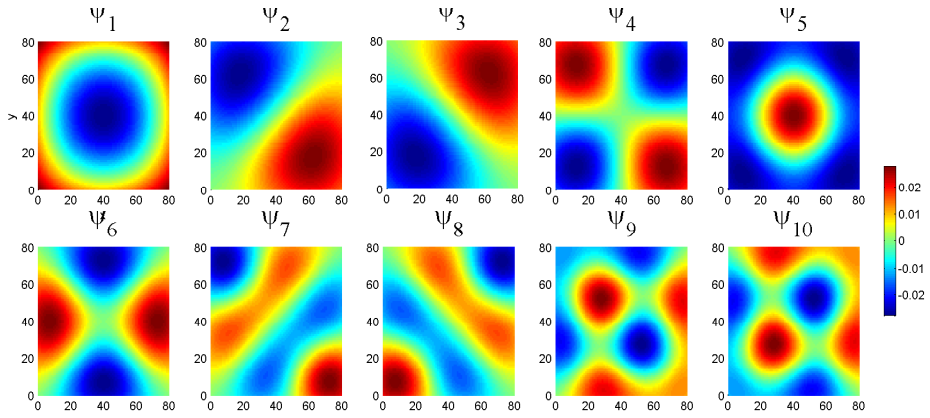

Figure 3.10: The eigenfunctions corresponding to the 10 largest eigenvalues

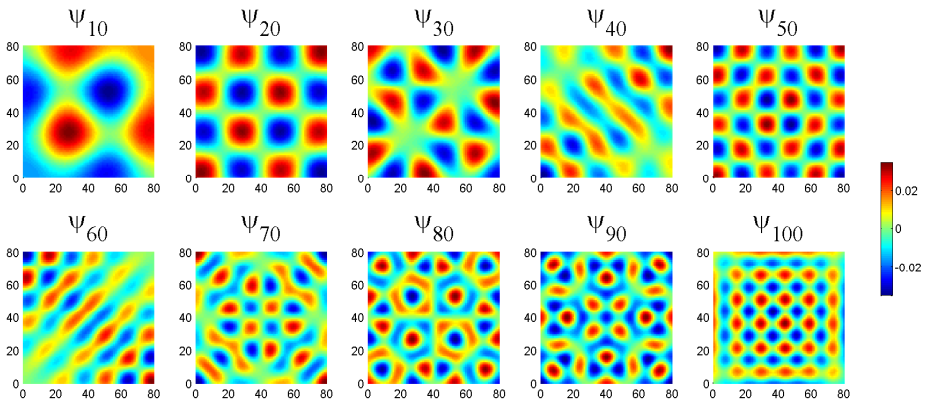

Figure 3.11: The high order eigenfunctions in the K-L expansion preserve the high frequency changes in the random field

3.3.2 Kernel Principal Component Analysis

The standard K-L expansion only preserves the two-point statistics of random fields. However, multi-point statistics have to be preserved in order to accurately model complex geological structures such as channels. Regardless of the number of eigenvalues retained, the K-L expansion serves as a linear approximation of a random field, and therefore will not capture nonlinear spatial relationships. One possible solution to the above problem is to map the realizations into a space at which the spatial relationships becomes approximately linear. Figure 3.12 demonstrate the basic idea behind nonlinear kernel principal component analysis [161, 154]. We note that many of the kernel PCA ideas and approaches described here are due to Scholkopf et al. [161] and have been described in detail previously within the machine learning literature. Sarma et al. [154] applied KPCA in the context of subsurface characterization and history matching. Our presentation in this section closely follows these earlier expositions.

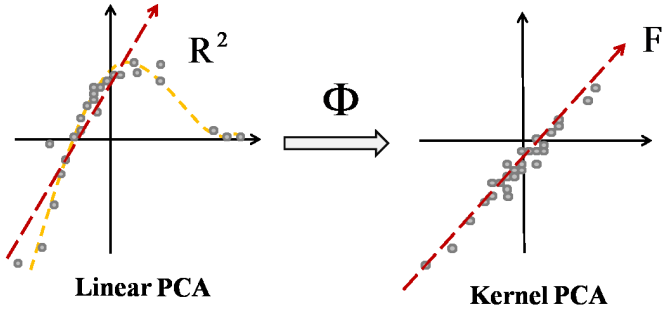

Figure 3.12: Basic idea behind kernel principal component analysis

Consider a nonlinear mapping Φ that transforms the input space R^{N_C} to another space F, called "the feature space", as:

$$\Phi : \quad R^{N_C} \rightarrow F; \qquad \mathbf{Y} = \Phi(y); \qquad y \in R^{N_C},\ \mathbf{Y} \in F. \tag{3.18}$$

Φ has to be chosen in a way that the realizations were nonlinearly correlated in the left graph of Figure 3.12, become approximated linearly correlated in the feature space F. Then the Karhunen-Loeve expansion can be applied in the feature space to reduce the dimension of the random field. To clarify further, in the context of geostatistical simulation, consider that a large number of realizations (each of length N_C) of a channelized permeability field have been obtained using some geostatistical technique. The input space is R^{N_C} , and the realizations can be thought of as points in this space. These points (realizations) are nonlinearly related in R^{N_C}; if they were linearly related, standard Karhunen-Loeve expansion (linear PCA) could be performed in R^{N_C} to obtain a channelized realization. Kernel PCA captures the nonlinear relationship between these realizations, thereby allowing us to form new (channelized) realizations that satisfy this nonlinear relationship. The advantage of the kernel PCA representation is that it provides a way to reduce the dimension in the feature space.

To perform the Karhunen-Loeve expansion in a high dimensional feature space F, the kernel eigenvalue problem must be solved. Accordingly, the ensemble covariance matrix in the feature space F can be written as [154]:

$$\tilde{\mathbf{C}} = \frac{1}{N_R} \sum_{j=1}^{N_R} \Phi(y_j)\Phi(y_j)^T. \tag{3.19}$$

where $\Phi(y_k)$ are the maps of realizations $y_k,\ k = 1, \ldots, N_R$ in the feature space and assumed to be centered (if not, they can be centered as in Scholkopf et al. [161]). The dimension of the ensemble covariance matrix in feature space is $N_F \times N_F$, where N_F is the length of $\Phi(y)$ and could be extremely large. Similar to the K-L expansion, to find the principal components in the feature space, the following eigenvalue problem must be solved:

$$\lambda \mathbf{v} = \tilde{\mathbf{C}}\mathbf{v}, \tag{3.20}$$

where λ is an eigenvalue of $\tilde{\mathbf{C}}$ and $\mathbf{v}$ is an eigenvector of $\tilde{\mathbf{C}}$.

Instead of solving the eigenvalue problem in the high-dimensional feature space F, similar to Section 3.3.1, a kernel eigenvalue problem associated with the equation (3.20) is solved:

$$N_R \lambda \alpha = \mathbf{K}\alpha. \tag{3.21}$$

Noting that here, the kernel matrix $\mathbf{K}$ is different from the kernel matrix applied for the eigenvalue problem in Section 3.3.1, and is defined as

$$\mathbf{K}: \quad K_{ij} = (\Phi(y_i).\Phi(y_j)). \tag{3.22}$$

The dimension of the kernel matrix is now $N_R \times N_R$, just as for the kernel eigenvalue problem in the previous section.

As in linear PCA, we can obtain all the non-zero eigenvalues λ and eigenvectors $\mathbf{v}$ of $\tilde{\mathbf{C}}$ from the eigenvalues $N_R\lambda$ and eigenvectors α of $\mathbf{K}$. There are only N_R non-zero eigenvalues of $\tilde{\mathbf{C}}$, if $N_R < N_F$; in general $N_R \ll N_F$.

Since for application of kernel PCA only the dot products in the feature space are required (and not $\Phi(y_k)$ itself), the kernel matrix can be calculated very efficiently by using a kernel function [161] that calculates the dot products directly from the realizations in the input space R^{N_C} as:

$$(\Phi(\mathbf{x}).\Phi(\mathbf{y})) = k(\mathbf{x}, \mathbf{y}). \tag{3.23}$$

Every kernel function that satisfies Mercer's condition [114], corresponds uniquely to a mapping Φ [161, 154]. Various kinds of kernel functions have been studied in the literature, but in this work, we mainly focus on the polynomial kernel defined as [161, 154]:

$$(\Phi(x).\Phi(y)) = k(x, y) = (x.y)^d, \tag{3.24}$$

where d is the order of the polynomial kernel. The polynomial kernel of order d, $(x.y)^d$, corresponds to a feature space F of d^{th} order monomials of R^{N_C}. For example, if we take $d = 2$ and the input space is R^3, we have [161]

$$\begin{aligned} k(\mathbf{x}, \mathbf{y}) &= (\mathbf{x}.\mathbf{y})^2; \qquad \mathbf{x} = (x_1, x_2, x_3)^T; \mathbf{y} = (y_1, y_2, y_3)^T, \\ k(\mathbf{x}, \mathbf{y}) &= x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2 + 2x_1x_2y_1y_2 + 2x_1x_3y_1y_3 + 2x_2x_3y_2y_3, \\ \Phi(\mathbf{x}) &= (x_1^2, x_2^2, x_3^2, \sqrt{2}x_1x_2, \sqrt{2}x_1x_3, \sqrt{2}x_2x_3)^T. \end{aligned} \tag{3.25}$$

Note that $\Phi(\mathbf{x})$ contains the product of 2 elements of $\mathbf{x}$ at a time and consequently the covariance matrix in the feature space F corresponds to fourth order or four-point statistics of the input space R^{N_C}. In general, for the polynomial kernel $(\mathbf{x}.\mathbf{y})^d$ of the order d, $\tilde{\mathbf{C}}$ corresponds to the $2d$th order moment of R^{N_C}. Hence, the Karhunen-Loeve expansion in the feature space F of the polynomial kernel $(\mathbf{x}.\mathbf{y})^d$ preserves the $2d^{th}$ order moment or $2d$-point statistics of R^{N_C}. Since we are interested not only in preserving the $2d^{th}$ order moment but all the moments up to the $2d^{th}$ order moment, the following kernel is used

$$(\Phi(\mathbf{x}).\Phi(\mathbf{y})) = k(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{d} (\mathbf{x}.\mathbf{y})^i. \tag{3.26}$$

A similar result can be obtained using an inhomogeneous polynomial kernel as described in Scholkopf [161]. In order to obtain an expression for random field $\mathbf{y}$ in the original space of R^{N_C}, an inverse $\mathbf{\Phi}$ map of $\mathbf{Y}$ is required as $\mathbf{y} = \mathbf{\Phi}^{-1}(\mathbf{Y})$. This is known as the pre-image problem [161, 154]. Since the dimension of the feature space F could be extremely large, the direct calculation of Φ^{-1} is practically impossible. To resolve this issue, the pre-image problem is posed as a minimization problem where $\mathbf{y}$ is sought in a way that the least-square error between $\Phi(\mathbf{y})$ and $\mathbf{Y}$ is minimized [161], details can be found at Sarma et. all [154].

In this work, we use kernel PCA as one of the several methods of non-linear dimensionality reduction for complex geological structures, where the standard Karhunen-Loeve expansion can not preserve the curvilinear patterns in a random field.

3.4 Polynomial Chaos Proxy

The use of polynomial chaos expansion discussed in Chapter 2 for uncertainty quantification was pioneered by Ghanem and Spanos [64] and has been successfully applied to various engineering problems [199, 8, 9, 110, 181, 198]. The first step in the application of polynomial chaos expansion as a proxy substitute for the reservoir simulation model is the representation of all the reservoir model input distributions in terms of a set of standardized random variables of zero mean and unit variance. The Karhunen-Loeve expansion or kernel PCA introduced in Section 3.3 can be employed to decompose the random field into a set of uncorrelated random variables. The second step is the derivation of orthogonal polynomials. This step is not required for the Gaussian random fields as the polynomial chaos expansions consist of Hermite polynomials. In case of non-Gaussian but analytical distributions like uniform distribution, Poisson distribution, etc., the correct set of orthogonal polynomials can be determined from the Askey scheme [199]. For empirical distributions, software such as ORTHOPOL can be used to derive the appropriate set of orthogonal polynomials [192]. Then the polynomial chaos expansion of the output random variable can be written in terms of the orthogonal polynomials derived in the previous step,

$$P(\vec{\xi}) = \sum_{j=0}^{m} \hat{a}_j \psi(\vec{\xi}). \tag{3.27}$$

For example, considering Gaussian random variables, the 1^{st} and 2^{nd} order polynomial chaos expansions are given as follows:

$$
\begin{aligned}
P(\xi_1, \xi_2, \ldots, \xi_N) &= a_{0,1} + \sum_{i=1}^{N} a_{i,1}\xi_i \\
P(\xi_1, \xi_2, \ldots, \xi_N) &= a_{0,2} + \sum_{i=1}^{N} a_{i,2}\xi_i + \sum_{i=1}^{N} a_{ii,2}(\xi_i^2 - 1) + \sum_{i=1}^{N-1}\sum_{j>i}^{N} a_{ij,2}\xi_i\xi_j.
\end{aligned}
\tag{3.28}
$$

Here N is the number of random variables used to represent the uncertainty in the model inputs, and the coefficients $a_{i,1}, a_{i,2}, a_{ii,2}, a_{ij,2} \quad \forall i = 1, \ldots N$ and $N > j > i$ are the coefficients to be estimated. However, the number of unknown coefficients to be determined increases quickly as the order is increased. For example, the number of unknown coefficients for the 2^{nd} and 3^{rd} order expansions are as follows:

$$
\begin{aligned}
M_2 &= 1 + 2N + \frac{N(N-1)}{2} \\
M_3 &= 1 + 3N + \frac{3N(N-1)}{2} + \frac{N(N-1)(N-2)}{6}.
\end{aligned}
\tag{3.29}
$$

The next step is the estimation of coefficients in the functional approximation. In order to find a good approximation for the model with the fewest number of full reservoir simulation runs, it is important to carefully select the parameters. As noted earlier in Section 2.4, there are various methods to estimate the unknown coefficients of the polynomial chaos expansions. Here in this chapter, we employ the regression-based PC), which uses a least-square technique to determine the polynomial chaos coefficients. The collocation points are selected from the roots of the next higher order orthogonal polynomial for each uncertain parameter. In general, the number of roots is usually much larger than the number of collocation points required, especially when the number of input random variables is large. The highest probability roots are usually chosen first as the collocation points. Once the M collocation points are chosen, the true model (full reservoir simulator) is evaluated M times. Refer to Chapter 2 for more detail on the implementation of the probabilistic collocation method.

Figure 3.13 illustrates the use of the polynomial chaos proxy along with the Karhunen-Loeve expansion. For the curvilinear random fields, the kernel PCA can be used along with the polynomial chaos proxy to give a better functional approximation.

3.5 History Matching with Semi-Definite Programming

The history matching problem includes the inversion of the observed production data (e.g., well pressures and flow rates) in order to determine reliable estimate of uncertain

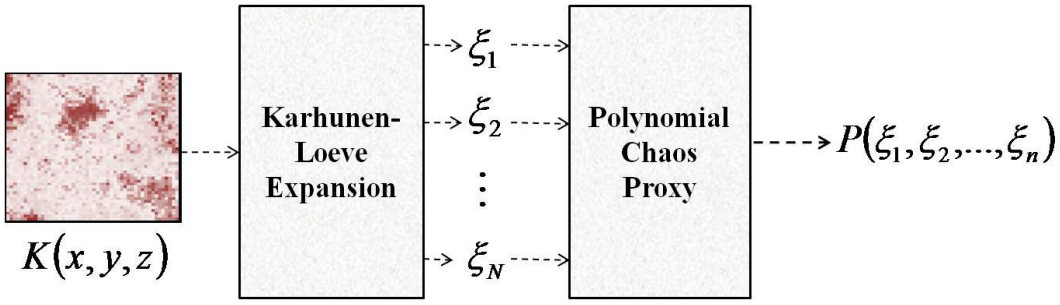

Figure 3.13: The polynomial chaos proxy is used along with the Karhunen-Loeve decomposition

model parameters (such as permeability). This problem can be written as a minimization problem of the misfit surface [176]. The misfit surface measures the difference between observed and modeled production data. If we assume the observations are independent and identically distributed, the misfit surface can be simply defined as:

$$\mathbf{S}(\xi) = \sum_{n=1}^{M} \frac{(obs_n - sim_n)^2}{2\sigma_n^2}, \tag{3.30}$$

where σ_n^2 is the standard deviation of the data (assuming independent data values here), and $n = 1, \ldots, M$ is the number of observations. The more general definition of the misfit surface will be presented later in the context of the Bayesian inference. If we plug the polynomial chaos expansion ($PC(\xi_1, \ldots, \xi_N)$) as a proxy substitute for the simulation model, we obtain the following expression for the misfit surface:

$$\mathbf{S}(\xi) = S_0 + \sum_{i=1}^{N} S_{i,1}\xi_i + \sum_{i=1}^{N} S_{ii,2}(\xi_i^2 - 1) + \sum_{i=1}^{N-1}\sum_{j>i}^{N} S_{ij,2}\xi_i\xi_j. \tag{3.31}$$

Then the history matching problem involves the minimization of a nonnegative multivariate polynomial. Minimizing a polynomial function subject to different types of constraints have been widely studied in the literature [71]. Because of its sheer size in the literature, it is not possible to even partially list the successful application of polynomial optimization in practice. Particularly, the search for efficient algorithms to find the global minimum of a multivariate polynomial function has been a priority for many mathematical programmers. Since evaluating a multivariate polynomial function is computationally cheap, stochastic algorithms can be applied directly to find the global minimum by running the polynomial model for possibly more than a million times. However, we use a much more efficient algorithm in this chapter, called "semi-definite programming" that transforms the polynomial optimization problem to

a sequence of convex optimization problems. Efficient solution methods for general convex optimization problems are well developed (Boyd and Vandenberghe [185]).

Shor [165] first showed the one-dimensional optimization of strictly-positive polynomials reduces to a convex problem. Lasserre [96] then demonstrated that the global minimization of sum of squares of multivariate polynomials can be approximated as closely as desired by solving a finite sequence of convex optimization problems of the same flavor as in the one-dimensional case. The crucial enabling fact is the computational tractability of the sum of squares decomposition for multivariate polynomials, coupled with powerful results from semi-algebraic geometry. Hence the polynomial optimization problem can be approximated as closely as desired with a convex optimization problem and accordingly can be efficiently solved. Several authors (see [121, 135, 97]) applied the same method of approximating the strictly-positive multivariate polynomials by convex (semi-definite) relaxations to efficiently solve the polynomial optimization problem.

In this work, we apply SDP to the misfit surface resulted from the polynomial chaos proxy, and solve for the global minimum. Since every reduced dimension vector of $\vec{\xi}$ corresponds to a large-scale model $K(x, y, z)$, we use an inverse map to obtain the high-resolution model which gives the misfit surface the global minimum.

There is one caveat here; the accuracy of the polynomial chaos proxy for a specific model depends on the possibility of that model. In the sense that the polynomial chaos expansion can capture the shape of the misfit function so long as the corresponding permeability models are not highly improbable. Hence, if the global minimum of the misfit surface computed by SDP corresponds to a highly improbable permeability model, the polynomial chaos proxy might not be accurate enough for this permeability model ($K_{min}(x, y, z)$). To ensure the desired accuracy, we need to run the full reservoir simulator for $K_{min}(x, y, z)$, and if the discrepancy of the result with the polynomial chaos proxy is considerable, the polynomial chaos proxy has to be trained with this permeability model. Then we repeat the procedure of the previous sections until the discrepancies become negligible.

3.6 Case Study

In this section, we apply the proposed history matching algorithm using the polynomial chaos proxy described above for the water flooding of a synthetic channel sand model. The simulation model represents a 2D horizontal square reservoir with two water injectors in the left corner and two producers in the right corner. Figure 3.14 illustrates

the configuration of the wells in the synthetic model. The reservoir covers an area of $500 \times 500\ m^2$ and has a thickness of $10\,m$. It is discretized into a 50×50 horizontal 2D grid. It is essentially an incompressible two-phase unit mobility oil-water system, with zero connate water saturation (initial water saturation) and zero oil residual system. The model is run for 1000 days with the injectors under rate control and the producers under bottom hole pressure (BHP) control with predefined rates and BHPs. The objective is to obtain the best estimate of the unknown permeability fields using observed data that entails the injector BHPs and producers the water cuts and oil flow rates. We note that an analogous system was considered previously by Sarma et al. [154] and was motivated by an earlier model due to Brouwer et al. [20].

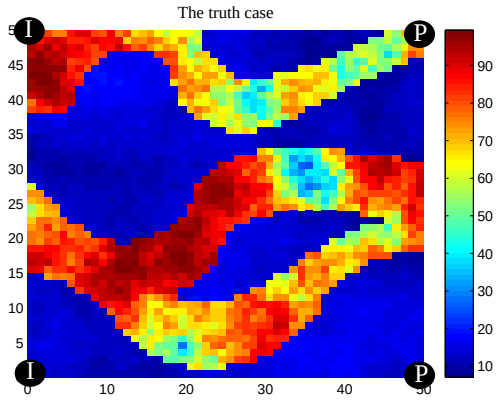

Figure 3.14: The synthetic water flooding model with two injectors on the left and two producers on the right. It also shows the reference permeability field used to generate the observed data of oil and water flow rates.

Although the permeability field is unknown, we assume that prior geological knowledge indicates that this is a fluvial channelized reservoir, which has a very fine sand texture of permeability about 80mD, and the background with permeability about 20mD. Fluvial is a term used in geology and it refers to the deposits and land-forms created by the processes associated with rivers and streams and it is explained as follows [117]:

'Rivers are perennial streams of water flowing in channels through valleys. Their flow is confined to the channels except during floods, when it overflows the riverbanks and covers parts of the adjacent floodplain. In high or hilly areas, rivers tend to flow in narrow, steep valleys, and thus the flow is very fast. Here the stream beds contain rounded boulders and cobbles and are very rough. On lower ground, rivers tend to flow more smoothly, although some stretches can have swift currents. Their stream beds contain smooth, rounded gravel and cobbles as well as sand and mud distributed according to the velocity and other characteristics of the stream-flow. River channels

tend to broaden and become less steep with distance downstream toward their junction with a master stream'.

Figure 3.15 shows the training image used for the prior model of the fluvial channelized reservoir under study.

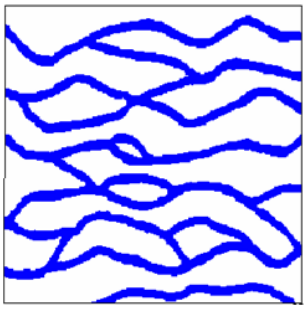

Figure 3.15: The training image for the fluvial channelized reservoir

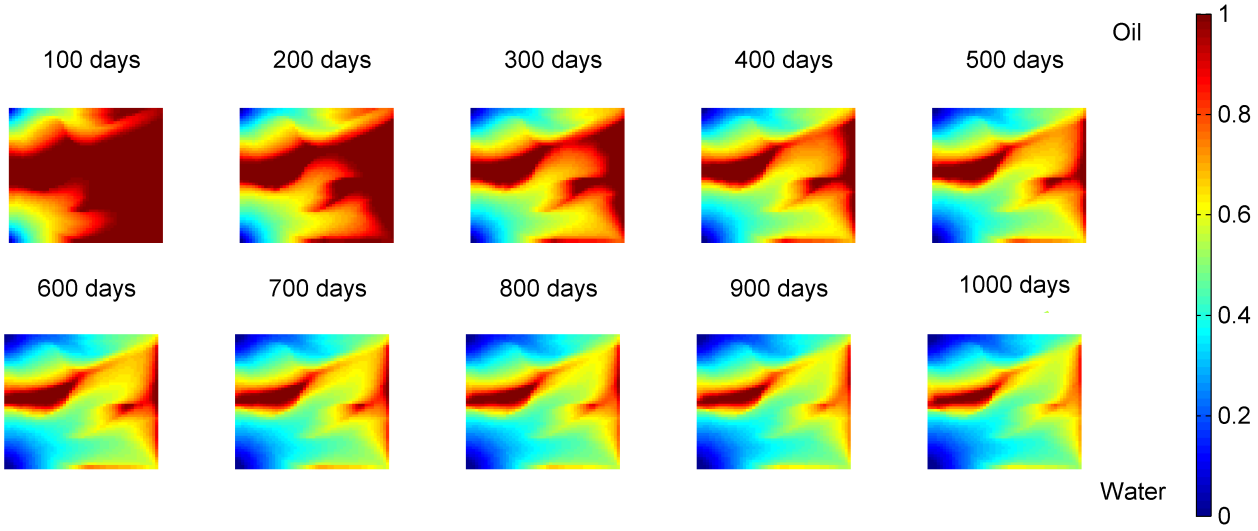

Figure 3.16: The saturation map for the water flooding process of the fluvial channelized reservoir, with two injectors placed in the left edge of the reservoir and two producers drilled in the right edge.

Figure 3.16 illustrates the water flooding process of the fluvial channel for 1000 days. The observed data includes only the history of water and oil production of two producers for 1000 days. The water injection rate is 500 $\frac{m^3}{day}$ for each injector, controlled by the rate. Figure 3.17 shows the observed data for each producer.

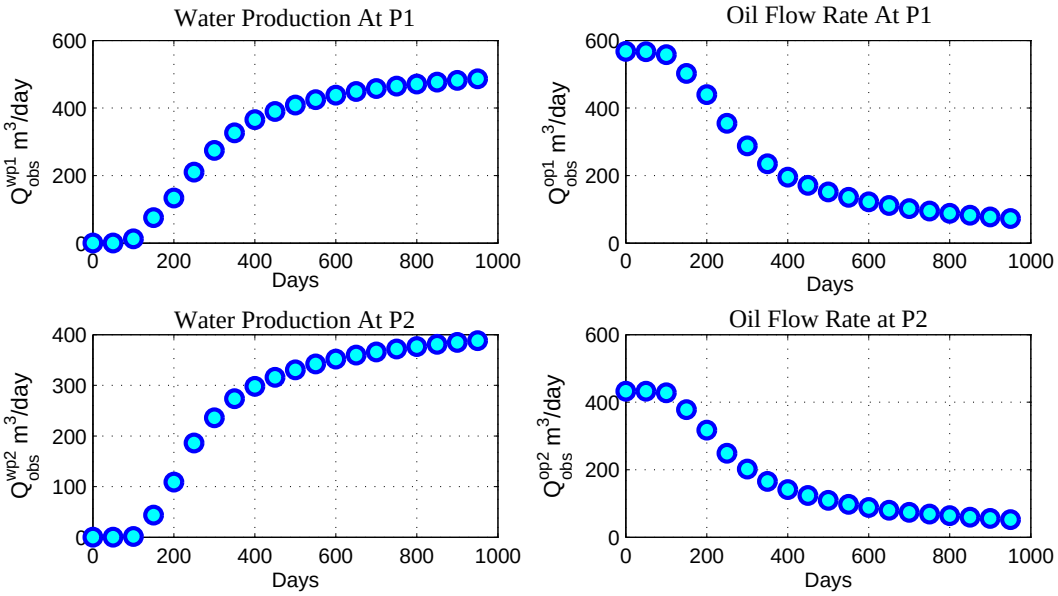

Figure 3.17: The observed history of the reservoir for each producer

The true permeability model which is used as a reference case to generate the production data is shown in Figure 3.14. The misfit surface is very complicated and has many local minimum points whose misfit values are close to zero. However, the high-order polynomial chaos expansion can capture this complex shape so long as the corresponding prior models are not highly improbable.

To find the permeability model which gives the global minimum to the misfit surface, we have to follow three main steps. First, it is necessary to reduce the dimension of the problem to be able to efficiently construct the polynomial chaos proxy. Second we need to train the polynomial chaos proxy and finally calculate the misfit surface and apply SDP to find the global minimum of it. The procedure of polynomial chaos proxy-based history matching is as follows:

- **Reduced-order parameterization of the permeability field**
  Using Petrel (Schlumberger) [160] with the training image of Figure 3.15, we produce 1000 realizations, and compute the ensemble covariance matrix. The training image is a heuristic reservoir analogue that imitates the geological heterogeneity and contiguous patterns of conductivity in the lithofacies under study. Training images are merely conceptual images to capture high-order statistics and they are not required to be constrained to any subsurface data. Based on the given training image of Figure 3.15, the probability model from which the value of a grid cell is generated, is specified. As a result, to obtain a realization constrained to

reservoir geological data, Petrel generates the value of one grid cell after another based on the conditional probability model inferred from the training image.

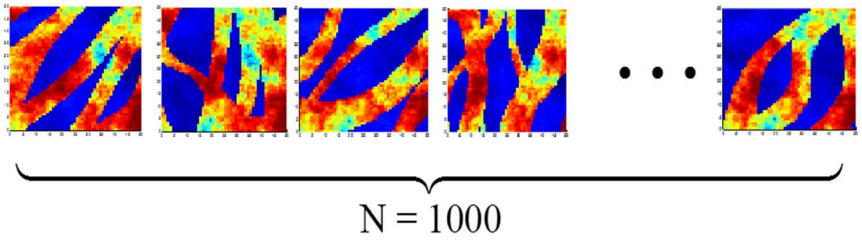

Figure 3.18: To train the polynomial chaos proxy, **N** realizations are generated based on the given prior distribution.

To reduce the dimension, we perform the Karhunen-Loeve transform on the ensemble covariance matrix of the N realizations, using the equation (3.5), and discard the smallest eigenvalues. Small eigenvalues incorporate the high frequency changes in the permeability field. Since in our problem the transition between the two different textures happens sharply (high frequency change), discarding the smallest eigenvalues will result in having a continuous transition from sand to shale. As far as our problem concerns the general shape of the fluvial channel, we neglect the high frequency changes and discard the small eigenvalues. To decide on the number of the eigenvalues retained, it has been shown that the

$$E(P) = \frac{\sum_{i=1}^{P} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \tag{3.32}$$

could be an effective metric to choose the optimal number of the eigenvalues [85]. For our problem, $E(10) = 0.56$ shows that 56% of the energy of the field can be captured using the first 10 eigenvalues and seems a reasonable number to retrieve the general shape of the channel. We solve the problem by preserving 10, 20, 30 and 40 eigenvalues ($P = 10, 20, 30, 40$) and compare the results, while $E(20) = 0.68, E(30) = 0.75, E(40) = 0.78$. We also employ the polynomial Kernel PCA retaining 10, 20, 30 and 40 eigenvalues in the feature space and compare the result to the linear PCA. It is worth noting that each vector in the feature space (with reduced dimension) corresponds to a high-resolution permeability field.

- **Construct the polynomial chaos proxy**
  To train the polynomial chaos proxy, we use regression-based PCM to compute the coefficients of the polynomial chaos expansion. We run the full reservoir simulator for each realization, and obtain the flow rates of oil and water for both of the

producers. Then we employ regression-based PCM to compute the coefficients of polynomial chaos expansion and achieve an expression for water and oil flow rates.

$$\begin{aligned}
\mathbf{Q^{op1}}(\xi,\mathbf{t}) &= Q_0^{op1} + \sum_{i=1}^{P} Q_{i,1}^{op1}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{op1}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{op2}}(\xi,\mathbf{t}) &= Q_0^{op2} + \sum_{i=1}^{P} Q_{i,1}^{op2}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op2}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{op2}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{wp1}}(\xi,\mathbf{t}) &= Q_0^{wp1} + \sum_{i=1}^{P} Q_{i,1}^{wp1}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{wp1}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{wp1}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{wp2}}(\xi,\mathbf{t}) &= Q_0^{wp2} + \sum_{i=1}^{P} Q_{i,1}^{wp2}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{wp2}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{wp2}(\xi_i\xi_j) + \dots .
\end{aligned} \tag{3.33}$$

As mentioned earlier in Section 2.4, in regression-based PCM the coefficients of polynomial chaos expansion are calculated in a way that minimizes the sum of squared errors. The error is defined as the difference between the full reservoir simulator result and the fitted value provided by the polynomial chaos proxy, for $\mathbf{N}$ training data. The method of least squares is a standard approach to the approximate solution of over-determined systems, i.e., sets of equations in which there are more equations than unknowns. Here the number of training simulation $\mathbf{N}$ is smaller than the number of polynomial chaos coefficients. Regression-based PCM ensures that the overall polynomial chaos model minimizes the sum of the squares of the errors made in the results of every single equation. For each realization $K_j(x,y)$ $(j = 1,\dots,N)$, we have four equations for $\mathbf{Q^j_{op1}}(\mathbf{t}), \mathbf{Q^j_{wp1}}(\mathbf{t}), \mathbf{Q^j_{op2}}(\mathbf{t}), \mathbf{Q^j_{wp2}}(\mathbf{t})$, where the unknowns are the polynomial chaos coefficients in the equation (3.33). Regression-based PCM adjusts the coefficients of polynomial chaos proxy to best fit the training data set.

To check the convergence of the polynomial chaos expansion, we generate few more realizations and compare the polynomial chaos results with the full reservoir simulator results to see if the proxy model is accurate enough. If the discrepancy between the results is large, it shows that we need more realizations to accurately train the polynomial chaos proxy.

- **Apply SDP to find the global minimum of the misfit surface:**
  Using the misfit definition of Section 3.5 as the difference between the observed and modeled production data, we have:

$$\begin{aligned}\mathbf{S}(\xi_{\mathbf{1}},\xi_{\mathbf{2}},\ldots,\xi_{\mathbf{P}}) = &\sum_{i=1}^{M}\frac{(\mathbf{Q}_{\mathbf{proxy}}^{\mathbf{op1}}-\mathbf{Q}_{\mathbf{obs}}^{\mathbf{op1}})_i^2}{\sigma_i^2}+\sum_{i=1}^{M}\frac{(\mathbf{Q}_{\mathbf{proxy}}^{\mathbf{op2}}-\mathbf{Q}_{\mathbf{obs}}^{\mathbf{op2}})_i^2}{\sigma^2}+\ldots\\ &+\sum_{i=1}^{M}\frac{(\mathbf{Q}_{\mathbf{proxy}}^{\mathbf{wp1}}-\mathbf{Q}_{\mathbf{obs}}^{\mathbf{wp1}})_i^2}{\sigma^2}+\sum_{i=1}^{M}\frac{(\mathbf{Q}_{\mathbf{proxy}}^{\mathbf{wp2}}-\mathbf{Q}_{\mathbf{obs}}^{\mathbf{wp2}})_i^2}{\sigma^2}.\end{aligned} \tag{3.34}$$

  If we plug the polynomial chaos expression of the equation (3.33) for water and oil flow rates of each producer in the equation (3.34), we obtain the following approximation for the misfit surface, in which the misfit coefficients $(S_0, S_{i,1}, S_{ii,2}, \ldots)$ are known:

$$\mathbf{S}(\xi_{\mathbf{1}},\xi_{\mathbf{2}},\ldots,\xi_{\mathbf{P}}) = S_0+\sum_{i=1}^{P}S_{i,1}\xi_i+\sum_{i=1}^{P}S_{ii,2}(\xi_i^2-1)+\sum_{i=1}^{P-1}\sum_{j>i}^{P}S_{ij,2}(\xi_i\xi_j)+\ldots \tag{3.35}$$

  Since the misfit surface is approximated by a non-negative multivariate polynomial, we can apply SDP to efficiently solve for the global minimum of the misfit surface. Every vector of ξ corresponds to a permeability field $K(x, y)$, consequently we can obtain the permeability field that gives the global minimum to the misfit surface.

Figure 3.19 shows the global minimum of misfit function. For this simulation the polynomial kernel PCA of order 3 is used to reduce the dimensions, while it retains the 40 largest eigenvalues and also the polynomial chaos expansion of order 12 is employed. The accuracy of history matching by the polynomial chaos proxy depends on the quality of the dimensionality-reduction, the order of the polynomial chaos expansion, the number of training runs and the prior distribution used to generate realizations. Figure 3.20 illustrates the effect of using two methods of dimensionality reduction (the Karhunen-Loeve expansion and kernel PCA) explained in Section 3.3, with different number of eigenvalues retained. It demonstrates that if we retain more eigenvalues in the feature space, the high-frequency changes are better preserved in the final answer.

To compare the accuracy of the estimation of the permeability field in Figure 3.20, we define the relative error of the estimation as:

$$\text{Relative error} = \frac{\parallel K_e - \mathbf{K}^{\dagger} \parallel^2}{\parallel \mathbf{K}^{\dagger} \parallel^2} \tag{3.36}$$

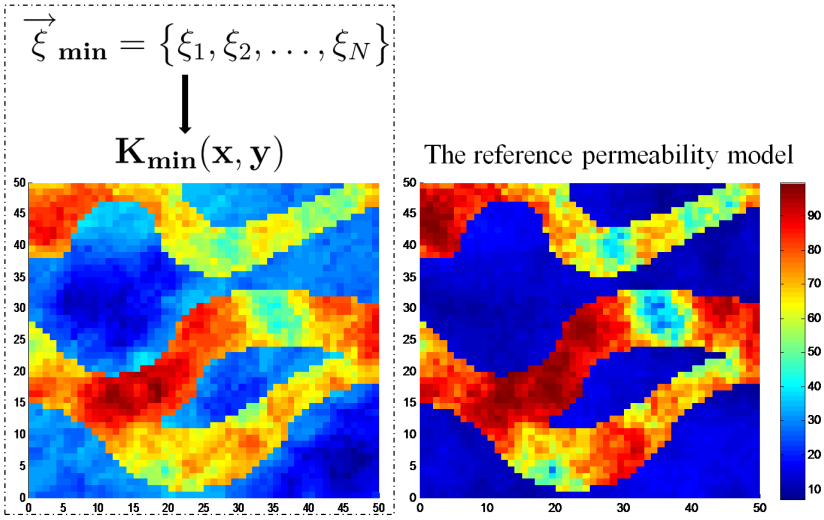

Figure 3.19: The global minimum of the misfit surface, 40 largest eigenvalues are retained in the feature space of the 3^{rd} order polynomial kernel PCA.

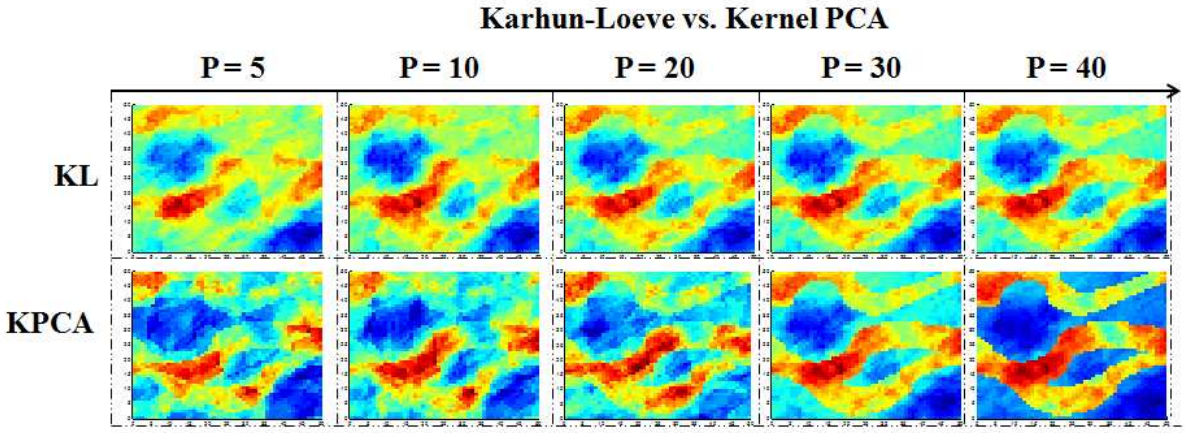

Figure 3.20: The comparison between the Karhunen-Loeve parameterization and kernel PCA for different number of dimensions in the reduced dimension space, while the polynomial chaos of order 10 is used as a surrogate model and the 3^{rd} order polynomial kernel used for KPCA results.

where K_e is the estimated permeability field and $\mathbf{K}^{\dagger}$ is the truth permeability case. Table 3.1 shows the relative error of the permeability field estimation for the results of Figure 3.20.

For this synthetic reservoir case study, the full reservoir simulator is implemented using MATLAB on a local workstation. The running time for a full reservoir simulation is $F = 185s$ while the pseudo-simulator using the polynomial chaos proxy has a running time of $\alpha = 3s$. The CPU time for each estimation of Figure 3.20 is shown in

Table 3.1: The relative error of the Karhunen-Loeve parameterization and Kernel PCA for different number of eigenvectors retained, while the polynomial chaos of order 10 is used as a surrogate model and the 3^{rd} order polynomial kernel used for KPCA results.

	P=5	P=10	P=20	P=30	P=40
KL	0.453	0.371	0.292	0.196	0.182
Kernel PCA	0.326	0.290	0.217	0.175	0.091

Table 3.2: The CPU time for the estimations of Figure 3.20 divided by the CPU time for a full reservoir simulation. The polynomial chaos of order 10 is used as a surrogate model and the 3^{rd} order polynomial kernel used for KPCA results.

	P=5	P=10	P=20	P=30	P=40
KL	1041	1048	1064	1081	1097
Kernel PCA	1443	1532	1709	1887	2064

Table 3.2, noting the fact that they are divided by F (the CPU time for a full reservoir simulation). It illustrates the equivalent number of full reservoir simulation required for each estimation.

Table 3.3 shows that higher order of the polynomial chaos expansion gives a more accurate permeability model. There is usually a trade-off between the number of training runs of the full reservoir simulator and the order of polynomial chaos expansion. As far as the global minimum is mainly concerned in this chapter, low orders of polynomial chaos expansion gives reasonable accuracy. Higher order of polynomial chaos proxy requires more trial runs to give more accurate approximations. For example, $N = 1000$ realizations are not enough to accurately compute the coefficients of the polynomial chaos expansion of order 12 with 40 random variables (40 eigenvalue is retained). To construct the polynomial chaos proxy with 40 random variables, we use the reduced-terms polynomial chaos expansion which only preserves the relevant terms. However, the polynomial chaos proxy when it is used along with the eigenfunction decomposition techniques, produce a highly sparse polynomial chaos expansion. Therefore, only few terms in the polynomial chaos representation are relevant and the rest could be discarded. This produces a reduced-terms polynomial chaos proxy that will be explained in detail in the next chapter. Accordingly, the high order (12^{th} order) reduced-terms polynomial chaos with 40 random variables can be obtained using $N = 1000$ realizations.

To find the accurate locations of local minima, we need to use higher orders of the polynomial chaos proxy. It is mainly due to the complex shape of the misfit surface. To better visualize this, let us compute the misfit surface with the full-reservoir simulator

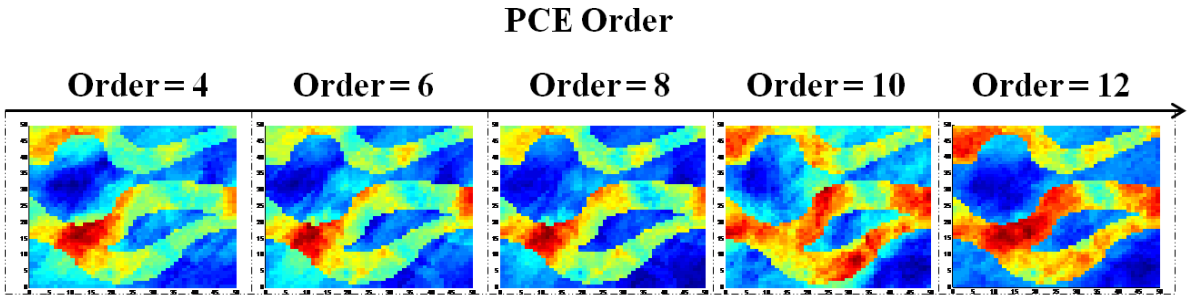

Figure 3.21: When the order of polynomial chaos expansion increases, the accuracy of the prediction improves. For the above comparison polynomial Kernel PCA of order 3 with 40 number of eigenvalues is used.

Table 3.3: The relative error of the polynomial chaos proxy-based estimation of the permeability field for different orders of the polynomial chaos expansion, while the polynomial Kernel PCA of order 3 with 40 number of eigenvalues is used.

	Order=4	Order=6	Order=8	Order=10	Order=12
Relative error	0.181	0.133	0.124	0.091	0.073

Table 3.4: The CPU time for the estimations of Figure 3.21 divided by the CPU time for a full reservoir simulation. The polynomial Kernel PCA of order 3 with 40 number of eigenvalues is used.

	Order=4	Order=6	Order=8	Order=10	Order=12
KL	1691	1824	1861	2064	2142

for different values of ξ_1 and ξ_2 (the random variables associated with the two largest eigenvalues of K-L expansion in the feature space), while other ξ_i's are kept at the maximum likelihood value. Figure 3.22 illustrates that even this reduced-dimension misfit surface, which is much simpler than the actual high-dimensional misfit surface, has many local minima.

The complexity of the reduced-dimension approximation of the misfit surface can be better illustrated in the contour map of Figure 3.23. The actual misfit surface is obviously much more involved. The low-order polynomial chaos proxy can give a reasonable approximation of the global minimum, but to probe the location of local minima, the higher-order terms have to be accurately computed.

Figure 3.24 suggests that the low-order (5^{th} order) polynomial chaos proxy can not capture the local minima accurately. However, the higher-order (15^{th} order) polynomial chaos proxy of Figure 3.26 gives a better approximation of the reduced-dimension misfit surface and accordingly the local minima. The grey surface shows the polynomial

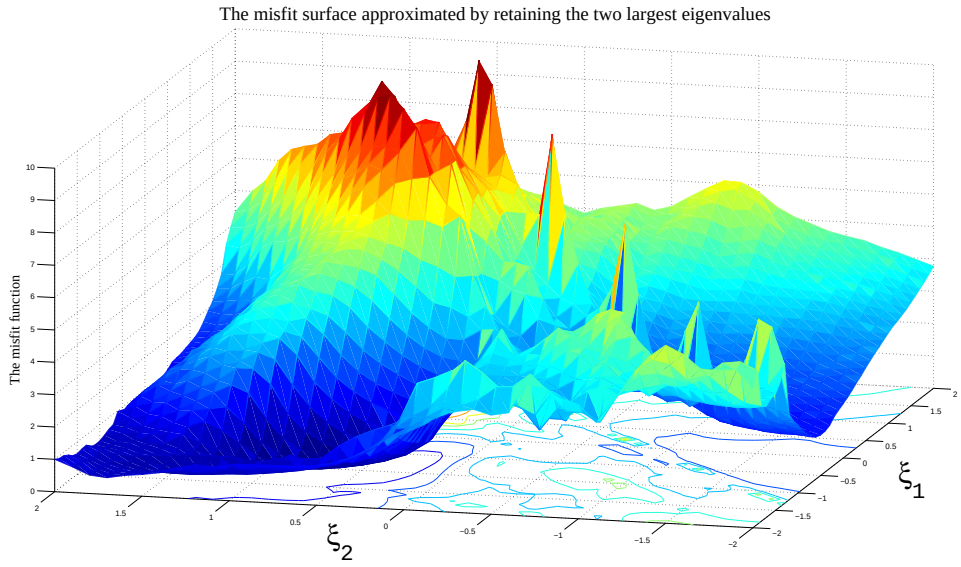

Figure 3.22: The reduced-dimension approximation of the misfit surface, the two largest eigenvalues are retained.

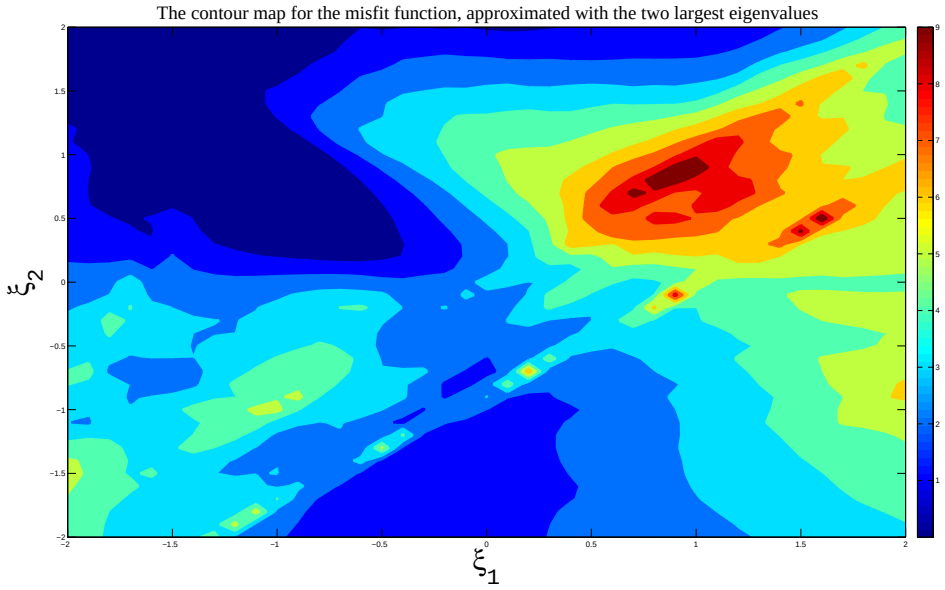

Figure 3.23: The contour map of the reduced-dimension misfit surface.

chaos proxy and the colored surface represents the reduced-dimension approximation of the misfit surface. The local minima can be accurately captured so long as their corresponding prior models are not highly improbable. To achieve high-order polynomial chaos expansion accurately, we have to increase the number of trial runs which reduces the computational efficiency of the algorithm.

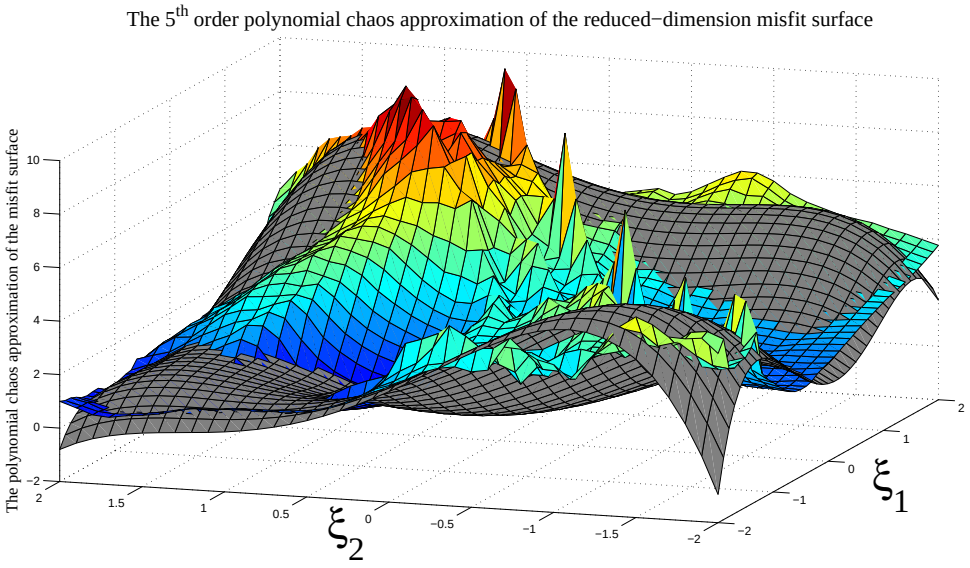

Figure 3.24: The 5^{th} order polynomial chaos approximation of the reduced-dimension misfit surface. The grey surface shows the 5^{th} order polynomial chaos and the colored surface represents the reduced-dimension approximation of the misfit surface.

The convergence of the algorithm depends on the quality of the given prior realizations. If the truth case lies in a highly improbable region of the prior distribution, the method may never converge to the solution. In this case study, the truth case almost lies a in a linear space of the given realizations, as Figure 3.20 suggests, and in the first iteration, the algorithm can approximately find the global minimum.

Tables 3.2 and 3.4 demonstrates the efficiency of the proposed algorithm as, for example, with the equivalent running time of 1080 full reservoir simulations, the relative error of 0.2 could be achieved, comparing to the current stochastic optimization algorithms that typically need hundreds of thousands of full reservoir simulations to approximately find the global minimum [126, 129]. In the next chapter we will show that for the similar problem, more than 10^5 full reservoir simulation runs are required for the stochastic algorithms to converge.

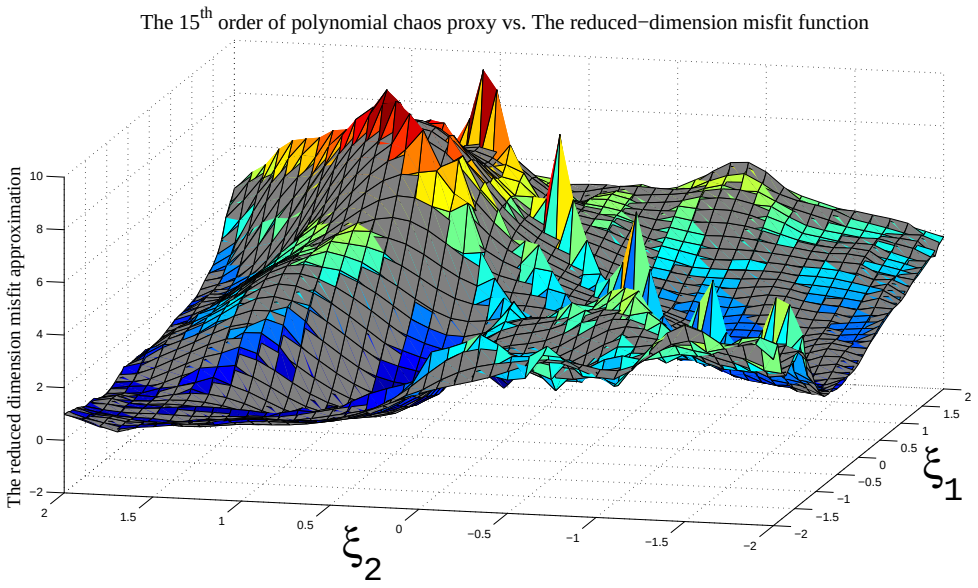

Figure 3.25: The 15^{th} order polynomial chaos approximation of the reduced-dimension misfit surface. The grey surface shows the 15^{th} order polynomial chaos and the colored surface represents the reduced-dimension approximation of the misfit surface.

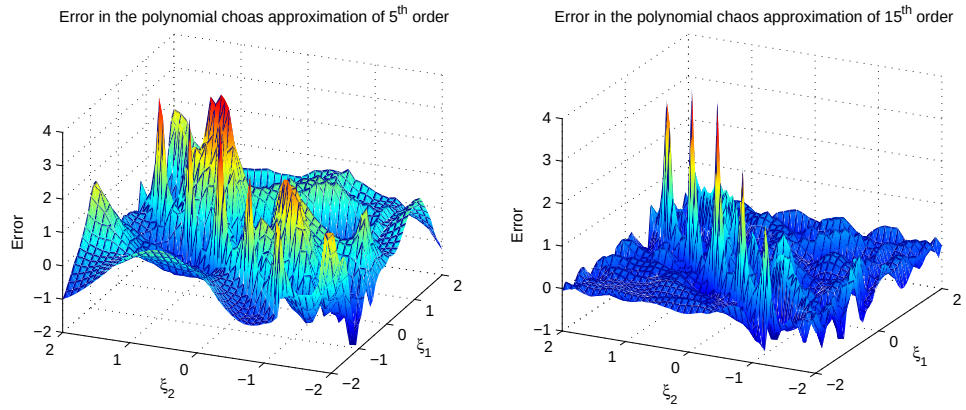

Figure 3.26: The error of the estimation of the reduced-dimension misfit surface using the polynomial chaos expansion of 5^{th} order (on the left) and the polynomial chaos expansion of 15^{th} order (on the right).

3.7 Summary

A novel polynomial chaos proxy-based automatic history matching is presented that can be employed to solve for the complex geological structures in inverse problems. For complex geological structures (characterized by multi-point Geostatistics), Kernel PCA is applied to reduce the dimension of the random field. We discussed that the quality of the reduced-order modeling lies in the connectivity (correlation) of geological random

fields. The more correlated a random field is, the more effectively it can be represented in a reduced dimension space. The reduced-order modeling is an essential part of the polynomial chaos proxy-based history matching as the curse of dimensionality reduces the efficiency of the surrogate model.

The choice of polynomial chaos expansion as the proxy substitute for full-reservoir simulator was proposed as it has the advantage over all other surrogate models that it guarantees the convergence in probability to output random variable. For Gaussian geological random fields, the polynomial chaos proxy with Hermite basis renders an exponential rate of convergence. For geological random fields with other distributions, the general polynomial chaos proxy framework was explained which adaptively chooses the orthogonal basis to have a faster rate of convergence.

The proxy-modeling methodology, also known as meta-modeling or surrogate modeling, is a computationally cheap alternative to full numerical simulation in assisted history matching, production optimization and forecasting has been studied by several authors. Different type of proxies and polynomials, e.g. Taylor series, have been successfully examined in history matching and production optimization. However, with increasing complexity of the solution space and number of uncertainties in the geological model, the accuracy of proxy-based history matching decreases. This is due to the fact that the aforementioned proxies are not able to probe the solution space systematically. But the accuracy of the polynomial chaos expansion is in accordance with the possibility of a geological model in the prior distribution.

By using the polynomial chaos proxy, the misfit surface is approximated by multivariate polynomials. As discussed, there are extremely efficient ways of minimizing a non-negative multivariate polynomials. Semi-definite programming is used as a new method of approximating the strictly-positive multivariate polynomials by convex (semi-definite) relaxations to efficiently find the global minimum of the misfit surface. The result is validated with the full-reservoir simulator and if the discrepancy is considerable, the polynomial chaos proxy is trained with the new geological model.

A simple case study was conducted to examine the applicability of the polynomial chaos proxy-based history matching. Although the misfit surface was very complex, the polynomial chaos proxy demonstrated to efficiently capture the global minimum of the misfit surface. Increasing the order of polynomial chaos expansion, the number of trial runs and the number of eigenvalues retained in the feature space was shown to improve the accuracy of the result.

The proposed general polynomial chaos proxy can be coherently used in other reservoir optimization work flows and uncertainty assessments.

Chapter 4

Efficient Markov Chain Monte Carlo Sampling Using Polynomial Chaos Expansion

4.1 Introduction

The traditional history matching usually obtains only a single history matched model, with uncertainty assessment provided by sensitivity calculations around the matched model. Modern reservoir management has a paramount focus on predicting the likely range of field recoveries and consequently providing economic evaluations of different field development strategies. It takes account of any errors in the observed history of the reservoir, and retrieves reservoir models whose simulation results lie within the vicinity of observed data, and uses them to estimate ranges in likely recovery factors. To achieve these reservoir models, there are two main approaches in the literature, one based on the optimization methods and the other based on the Bayesian inference.

The optimization methods change unknown parameter values through an automated process to obtain reservoir models within the allowed range of misfit. Various optimization techniques have been developed in the literature, including Genetic Algorithms [27], Particle Swarm Optimization [139], Neighborhood Algorithm [151], Estimation of Distribution [138], Levenberg-Marquardt [101] and LBFGS [202]. There is a recent review paper on history matching [126] which presents a comprehensive list of references on the optimization methods applied for history matching and uncertainty quantification. These algorithms generate a sequence of parameter values that generally improve the

history match as time evolves. Existing optimization methods can be roughly classified into two general categories: the stochastic algorithms and the gradient-based methods. The gradient-based algorithms have several inherent limitations, including the need to compute the gradients at each step of the optimization process. A definite advantage of stochastic algorithms is that they are able to easily honor complex geological constraints by preserving multipoint statistics present in the prior geological model; the main drawback of these approaches is their inefficiency, as they require large number of simulations for convergence [197, 103]. However, most of these optimization-based algorithms do not provide any statistically valid method of assessing uncertainty without additional calculations. For example, Genetic Algorithms [27] and Particle Swarm Optimization [139] do not correspond to a valid sampling mechanism. The reason for this is that the distribution of parameter values is mainly controlled by the algorithm settings [47]. This needs to be corrected by running a second code to compute probabilities associated with each set of parameters [151].

Approaches based on the Bayesian inference, on the other hand, aim at estimating the posterior probability for the reservoir properties [81]. Existing Bayesian inference methods broadly entails algorithms based on particle filters such as the Ensemble Kalman Filter (EnKF) [49, 1] and the Markov Chain Monte Carlo (MCMC) approaches [127, 104].

MCMC methods are often used to probe the posterior probability distribution in the Bayesian inference inverse problems. Many MCMC methods move around the target distribution in relatively small steps, with no tendency for the steps to proceed in the same direction [176]. Among these methods are the Gibbs sampling method, the Metropolis-Hasting algorithms and the slice sampling algorithm [176]. These methods are easy to implement and analyze, but unfortunately it can take a long time for the walker to explore all of the space. The walker will often double back and cover ground already covered [176]. The difficult problem is to determine how many steps are needed to converge to the stationary distribution within an acceptable error. A good chain will have a rapid mixing at which the stationary distribution is reached quickly starting from an arbitrary position. Variants of MCMC techniques have been developed in the literature to increase the convergence rate to the target distribution, but they are usually hard to implement [176]. Among these methods are Langevin MCMC [172], Hamiltonian Monte Carlo [42] and combinations of evolutionary algorithms with MCMC [74, 186].

Oliver et al. [127] is one of the pioneer works on using MCMC in the context of reservoir simulation. He used MCMC methods for conditioning a permeability field

to pressure data. Ma et. al [104] proposed to use MCMC (a two-stage approach) for conditioning permeability fields. More recently, Emerick and Reynolds [45] proposed to use MCMC to improve the sampling obtained by the ENKF method.

However, typical uses of MCMC methods need more than 10^5 steps to sample from the target distribution with a reasonable error [129]. For reservoir studies, they require large number of simulations for convergence to reach the equilibrium state and it is practically infeasible. Hence, the main disadvantage of these approaches is their inefficiency. Therefore, it can be extremely time-consuming if high resolution models are used. This is particularly of concern in closed-loop reservoir management, which requires continuous real-time use of history matching and uncertainty quantification algorithms [78, 155]. Thus, there is a significant need for an efficient proxy (or surrogate) model that can predict simulation results with a reasonable accuracy.

The choice of the polynomial chaos expansions as the proxy substitute for full-reservoir simulator was discussed in detail in Chapter 3. It was explained that the polynomial chaos proxy has the advantage over all other surrogate models that it systematically guarantees the convergence in probability to output random variable. For the Gaussian random fields, the polynomial chaos proxy with Hermite basis renders an exponential rate of convergence and it is very efficient. For the non-Gaussian random fields, the general polynomial chaos proxy framework was explained which adaptively chooses the orthogonal basis to have a faster rate of convergence.

In this chapter, we propose the application of polynomial chaos expansions as a proxy substitute for the full reservoir simulator proxy when applied with the MCMC method to efficiently sample from the posterior probability density function of reservoir random parameters. Figure 4.1 illustrates the framework for the polynomial chaos proxy-based Bayesian inference using the MCMC method. It includes four main components; the dimensionality reduction component, the polynomial chaos proxy, the Bayesian inference formulation and the MCMC component. The first two components were discussed in detail in Chapter 3. We use PCA (and also KPCA to preserve the multipoint statistics) to decompose the geological parameters into a lower dimension in the feature space. The polynomial chaos proxy is then trained with the reduced-order parameters of reservoir model. The Bayesian inference provides a mathematical formulation for the posterior distribution of reservoir parameters. Instead of running the full reservoir simulation, we use the polynomial chaos proxy for the Bayesian inference. Then, we apply the MCMC method to sample from the posterior distribution.

This chapter is structured as follows; first we review a summary of the Bayesian inference formulation for the posterior probability distribution. Then we explain the

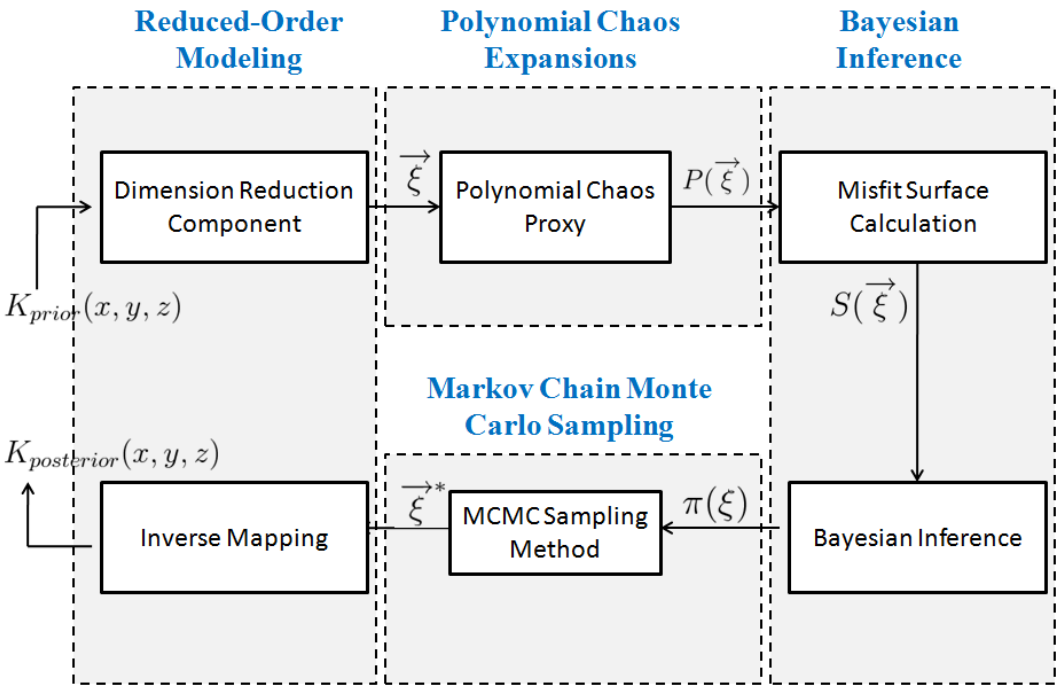

Figure 4.1: The framework for the MCMC method of Bayesian inference with polynomial chaos proxy

MCMC algorithm as a sampling method to obtain realizations from the posterior probability distribution. Similar to Chapter 3, the proposed algorithm has been examined on a two-dimensional synthetic case of a fluvial channelized reservoir. We note that our presentation here in the Bayesian inference section and the MCMC algorithms closely follows the exposition of Tarantola (2005) [176].

4.2 Bayesian Inference

Let Ω describes a physical system in a wide sense. In the way that it consists of a physical system plus measuring tools. For convenience of description, we assume Ω is discrete (or it has been discretized) and also parameterizable. Let $\mathbf{X} = (\mathbf{D}, \mathbf{P})$ denote the set of parameters describing Ω, where $\mathbf{D} = (D_1, \ldots, D_r)^T$ are data that contains the output of measuring instruments and $\mathbf{P} = (P_1, \ldots, P_s)^T$ express the physical parameters. Hence, any probability density function on $\mathbf{X}$ can be represented as:

$$f(x) = f(d, p). \tag{4.1}$$

In particular, we assume the existence of a *priori* probability density function

$$\rho(x) = \rho(d, p), \tag{4.2}$$

that represents the result of measurements and all a priori information of the physical parameters.

The physical systems are commonly assumed to be governed by a theoretical relationship between the values of the parameters:

$$\mathbf{d} = \mathbf{G}(\mathbf{p}). \tag{4.3}$$

where $d \in \mathbb{R}^r$ and $p \in \mathbb{R}^s$. This assumption is too restrictive, as in most cases for the forward problem, even if the value of $\mathbf{p}$ is given, the corresponding output of the system can not be exactly computed, for a host of reasons. We highlight two of these, particulary relevant to our discussion here:

- Our theory may be incomplete or include some random parameters
- The system may have been roughly parameterized.

Thinking probabilistically enables us to overcome these difficulties. In a sense that we assume not the exact relationship $\mathbf{d} = \mathbf{G}(\mathbf{p})$ governs the system but the probability density function for $\mathbf{d}$ given $\mathbf{p}$, i.e. the conditional probability density $\theta(\mathbf{d}|\mathbf{p})$ (see Figure 4.2). In general, we assume any theoretical relationship between parameters can be expressed as a joint probability density function:

$$\theta(x) = \theta(\mathbf{d}, \mathbf{p}). \tag{4.4}$$

From the definition of conditional probability [176], the joint probability density can be expressed as:

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p})\, \theta_{\mathbf{P}}(p), \tag{4.5}$$

where $\theta_{\mathbf{P}}(p)$ is the marginal probability density function for $\mathbf{P}$.

For the problems that the simplification of the equation (4.3) can be used, the theory does not impose any constraint on $\mathbf{P}$ but only in $\mathbf{D}$. Therefore, the equation (4.5) can be written as

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p})\, \mu_{\mathbf{P}}(p), \tag{4.6}$$

where $\mu(p)$ which represents the null information density (the state of total ignorance) [80, 148]. Usually the state of total ignorance corresponds to a uniform function $\mu(p)$ =const., sometimes it will not (see Tarantola [176] for more detail). The conditional probability distribution for the particular case of an exact theory clearly corresponds to $\theta(\mathbf{d}|\mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p}))$ where δ is the Dirac distribution. Hence, for the conditional probability density of an exact theory we obtain:

$$\theta(\mathbf{d}, \mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p}))\, \mu_{\mathbf{P}}(p). \tag{4.7}$$

In the class of problems where a rigorous computation of $\theta(\mathbf{d}|\mathbf{p})$ can not be made, but we have a heuristic idea of the theoretical error bound $\sigma_{\mathbf{T}}$, the following model is commonly used for the conditional density function [176]:

$$\theta(\mathbf{d}|\mathbf{p}) = \text{const. } \exp\Big\{ -\frac{1}{2}\frac{\| \mathbf{d} - \mathbf{G}(\mathbf{p}) \|^2}{\sigma_{\mathbf{T}}^2}\Big\}. \tag{4.8}$$

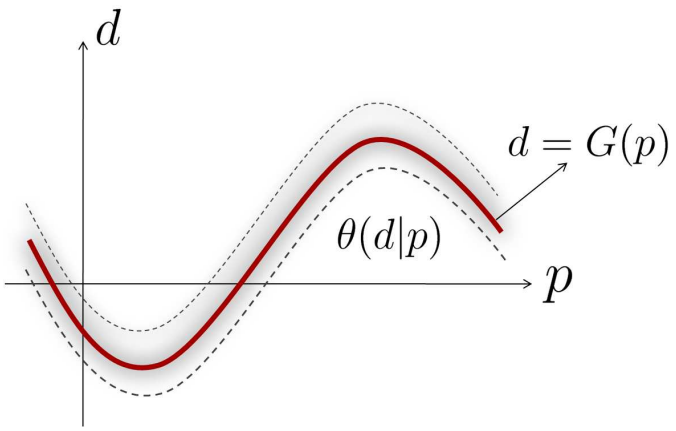

Figure 4.2: In the probabilistic framework, the relationship between system parameters is formulated with a probability density function $\theta(\mathbf{d}|\mathbf{p})$ rather than an exact functional relationship $\mathbf{d} = \mathbf{G}(\mathbf{p})$

The conjunction of the prior distribution $\rho(x)$ and the joint probability distribution $\theta(x)$ of the equation (4.4) brings a new state of information, named the *a posteriori information*. In the Bayesian statistics, the posterior probability distribution of a random event is the conditional probability that is assigned after all the relevant information and observations are taken into account. The joint posterior probability density function $\sigma(\mathbf{d}, \mathbf{p})$ is calculated using Bayes' theorem by [119]

$$\sigma(\mathbf{d}, \mathbf{p}) = \frac{\rho(\mathbf{d}, \mathbf{p})\theta(\mathbf{d}, \mathbf{p})}{\mu(\mathbf{d}, \mathbf{p})}. \tag{4.9}$$

Accordingly, the a posteriori marginal density function for the physical parameters can be computed as:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \int \frac{\rho(\mathbf{d}, \mathbf{p})\theta(\mathbf{d}, \mathbf{p})}{\mu(\mathbf{d}, \mathbf{p})} d\mathbf{d}. \tag{4.10}$$

The equation (4.10) transfers the information contained in the data set to the physical parameters using theoretical correlations that is probabilistically described by $\theta(\mathbf{d}, \mathbf{p})$.

In most cases the a priori information on $\mathbf{D}$ is independent from the a priori information on $\mathbf{P}$,

$$\rho(\mathbf{d}, \mathbf{p}) = \rho_{\mathbf{d}}(d)\, \rho_{\mathbf{P}}(p), \tag{4.11}$$

and the theoretical joint probability density function is obtained in the form of a conditional density function and the null information density function:

$$\theta(\mathbf{d}, \mathbf{p}) = \theta(\mathbf{d}|\mathbf{p})\, \mu_{\mathbf{P}}(p). \tag{4.12}$$

Therefore, the equation (4.10) can be simplified to:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \rho_{\mathbf{P}}(\mathbf{p}) \int \frac{\rho_{\mathbf{d}}(\mathbf{d})\theta(\mathbf{d}|\mathbf{p})}{\mu_{\mathbf{d}}(\mathbf{d})} d\mathbf{d}. \tag{4.13}$$

Conventionally the equation (4.13) is written as:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = \rho_{\mathbf{P}}(\mathbf{p})\mathbf{L}(\mathbf{p}), \tag{4.14}$$

where $\mathbf{L}(\mathbf{p})$ is the *likelihood function*,

$$\mathbf{L}(\mathbf{p}) = \int \frac{\rho_{\mathbf{d}}(\mathbf{d})\theta(\mathbf{d}|\mathbf{p})}{\mu_{\mathbf{d}}(\mathbf{d})} d\mathbf{d}, \tag{4.15}$$

which gives a measure of how good a model $\mathbf{p}$ is in explaining the data.

The equation (4.13) gives the solution of the general inverse problem. From $\sigma_{\mathbf{P}}(\mathbf{p})$ one can compute the mean values of the physical parameters, the median values, the maximum likelihood value or other statistical properties. Particularly, the probability of a model to satisfy some characteristics can be computed from $\sigma_{\mathbf{P}}(\mathbf{p})$ by integrating the probability density of the physical parameters over the region that satisfies the given characteristics. Additionally, different samples of the physical parameters can be drawn from $\sigma_{\mathbf{P}}(\mathbf{p})$ to provide a better understanding of a physical system.

The solution for an inverse problem exists simply when $\sigma_{\mathbf{P}}(\mathbf{p})$ as defined by the equation (4.13) does not correspond to the null density function. If the solution for $\sigma_{\mathbf{P}}(\mathbf{p})$ does not exist, it shows the incompatibility of the observed data, the prior assumption for the physical parameters or the theoretical information. The uniqueness of the solution for an inverse problem is evident, as by solution we mean the probability density function $\sigma_{\mathbf{P}}(\mathbf{p})$ itself.

In the following, we study the posterior distribution $\sigma_{\mathbf{P}}(\mathbf{p})$ for three different cases of modelization and observational uncertainties:

- **Negligible modelization uncertainties:**
  For an exact theory assumption $\theta(\mathbf{d}|\mathbf{p}) = \delta(\mathbf{d} - \mathbf{G}(\mathbf{p}))$ (or when modelization uncertainties are negligible compared to observational uncertainties), if the data space $\mathbf{D}$ is a linear space ($\mu_{\mathbf{D}}(\mathbf{d}) = \text{const.}$), we obtain:

  $$\sigma_{\mathbf{p}}(\mathbf{p}) = k\rho_{\mathbf{P}}(\mathbf{p})\rho_{\mathbf{D}}(\mathbf{G}(\mathbf{p})) \tag{4.16}$$

  where k is the normalization constant.

- **Negligible observational uncertainties:**
  Let $\mathbf{d_{obs}}$ denote the observed data values, the assumption of negligible observational uncertainties (with respect to modelization uncertainties) implies that:

$$\rho_{\mathbf{D}}(\mathbf{d}) = \delta(\mathbf{d} - \mathbf{d_{obs}}). \tag{4.17}$$

  Assuming the data space is a linear space , the equation (4.13) then gives:

$$\sigma_{\mathbf{p}}(\mathbf{p}) = k\rho_{\mathbf{P}}(\mathbf{p})\,\theta(\mathbf{d_{obs}}|\mathbf{p}). \tag{4.18}$$

- **The Gaussian modelization and observational uncertainties:**
  The Gaussian modelization corresponds to the assumption that the conditional probability distribution can be represented as:

$$\theta(\mathbf{d}|\mathbf{p}) = k_1 \exp\Big(-\frac{1}{2}(\mathbf{d} - \mathbf{G}(\mathbf{p}))^{\mathrm{T}}\mathbf{C_T^{-1}}(\mathbf{d} - \mathbf{G}(\mathbf{p}))\Big). \tag{4.19}$$

  Similarly, the Gaussian assumption for observational uncertainties implies that

$$\rho_{\mathbf{D}}(\mathbf{d}) = k_2 \exp\Big(-\frac{1}{2}((\mathbf{d} - \mathbf{d_{obs}})^{\mathrm{T}}\mathbf{C_d^{-1}}(\mathbf{d} - \mathbf{d_{obs}}))\Big), \tag{4.20}$$

  where k_1 and k_2 are the normalization constant. We use the notation $\mathbf{C_d}$ for the covariance matrix representing the measurement uncertainties and $\mathbf{C_T}$ for the covariance matrix representing the modelization uncertainties. Tarantola [176] demonstrated that for the posterior probability density function, equation (4.13) then gives

$$\sigma_{\mathbf{P}}(\mathbf{p}) = k\,\rho_{\mathbf{P}}(\mathbf{p})\exp\Big(-\frac{1}{2}((\mathbf{G}(\mathbf{P}) - \mathbf{d_{obs}})^{\mathrm{T}}\mathbf{C_D^{-1}}(\mathbf{G}(\mathbf{P}) - \mathbf{d_{obs}}))\Big), \tag{4.21}$$

  where

$$\mathbf{C_D} = \mathbf{C_d} + \mathbf{C_T}, \tag{4.22}$$

  and k is the normalization constant. This result is important because it shows that, in the Gaussian assumption, observational uncertainties and modelization uncertainties combine by addition of the respective covariance operators, even when the forward problem is nonlinear.

  We can write the equation (4.21) as

$$\boxed{\sigma_{\mathbf{P}}(\mathbf{p}) = k\,\rho_{\mathbf{P}}(\mathbf{p})\exp\Big(-\mathbf{S}(\mathbf{p})\Big)} \tag{4.23}$$

where the misfit function $\mathbf{S}(\mathbf{p})$ is the sum of squares:

$$\mathbf{S}(\mathbf{p}) = -\frac{1}{2}\Big((\mathbf{G}(\mathbf{p}) - \mathbf{d_{obs}})^T \mathbf{C_D^{-1}} (\mathbf{G}(\mathbf{p}) - \mathbf{d_{obs}})\Big). \tag{4.24}$$

If we further assume that the observations are independently identically distributed, for the misfit function of the equation (4.24) we obtain:

$$\mathbf{S}(\mathbf{p}) = -\frac{1}{2}\Big(\sum_i \frac{(\mathbf{G^i}(\mathbf{p}) - \mathbf{d^i_{obs}})^2}{\sigma_\mathbf{i}^\mathbf{2}}\Big), \tag{4.25}$$

which we used in Chapter 3 for the history matching problem.

Here in this work, we use the Gaussian modelization and assume that observations are also independently identically distributed. Recalling the geological parameterization, the dimensionality reduction and the polynomial chaos proxy component in Figure 4.3, let $\xi_1, \xi_2, \ldots, \xi_N$ be the reduced dimension parameters of the geological parameters $\mathbf{P}$, and the polynomial chaos proxy $\mathbf{PC}(\xi)$ be substituted for $\mathbf{G}(\mathbf{p})$ in the equation (4.25), we obtain the following expression for the misfit function

$$\mathbf{S}(\xi) = S_0 + \sum_{i=1}^{N} S_{i,1}\xi_i + \sum_{i=1}^{N} S_{ii,2}(\xi_i^2 - 1) + \ldots + \sum_{i_1,i_2,\ldots,i_N} S_{i_1,i_2,\ldots,i_N,n} H_n(\xi_1, \xi_2, \ldots, \xi_N) \tag{4.26}$$

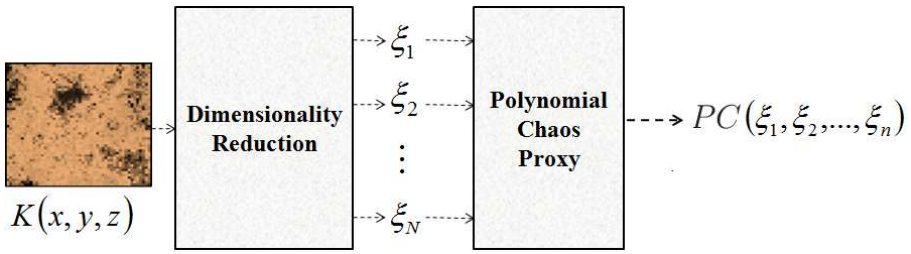

Figure 4.3: The exact model $\mathbf{G}(\mathbf{p})$ is replaced by the polynomial chaos proxy $\mathbf{PC}(\xi)$.

Consequently, the solution to the inverse problem of the system described above, corresponds to calculate the posterior distribution of the reduced dimension parameters $\xi_1, \xi_2, \ldots, \xi_N$. By replacing $\mathbf{S}(\xi)$ in the equation (4.23) with the misfit expansion of the equation (4.26), we can achieve an analytical expression for the posterior distribution of the reduced-dimension geological parameters

$$\sigma_\xi(\xi) = k\, \rho_\xi(\xi)\, \mathbf{exp}\{-(S_0 + \sum_{i=1}^{N} S_{i,1}\xi_i + \ldots + \sum_{i_1,i_2,\ldots,i_N} S_{i_1,i_2,\ldots,i_N,n} H_n(\xi_1, \xi_2, \ldots, \xi_N))\}, \tag{4.27}$$

where k is the normalization constant and $\rho_\xi(\xi)$ is the prior distribution of the reduced dimension geological parameters. By using the polynomial chaos proxy substitute for $\mathbf{G}(\mathbf{p})$, we hereby achieve an analytical expression for the posterior distribution. We can readily use the analytical formulation of the equation (4.27) for post processing (posterior mean, covariance, P_{10}, P_{90}, etc).

Since the general expression for the posterior distribution is nonlinear and also not necessarily Gaussian, we need to apply the MCMC methods to sample from the posterior distribution. In the next section we briefly introduce the fundamental steps in the MCMC sampling method.

4.3 Markov Chain Monte Carlo

In the previous section we demonstrated that the solution to an inverse problem is given by the posterior probability density function $\sigma_{\mathbf{P}}(\mathbf{p})$ that equals to the normalized product of the prior probability density function $\rho_{\mathbf{P}}(\mathbf{p})$ times a likelihood function $\mathbf{L}(\mathbf{p})$. The likelihood function is a measure of how good a model is in fitting the data. We also obtained an analytical approximation for posterior probability density function $\sigma_{\mathbf{P}}(\mathbf{p})$ by replacing the full reservoir simulator with the polynomial chaos proxy in the equation (4.27).

A major limitation towards the practical implementation of the Bayesian approaches in uncertainty quantification is that calculating the probability distribution $\sigma_{\mathbf{P}}(\mathbf{p})$ often involves the integration of high-dimensional functions [7]. This can be computationally infeasible. However, having a reasonable number of samples from a probability distribution allows us to quantify the uncertainty associated with a random variable using simple statistics. MCMC methods attempt to draw samples from some complex distribution of interest. Monte Carlo literally refers to random sampling and Markov chain restrict it to a sampling sequence that the probability of the next state, solely depends on the current state [7]. The state of the chain after a large number of steps that the Markov chain reaches its equilibrium distribution is then used as a sample from the desired distribution. The quality of samples generally improve as the number of steps increases. To conduct this Markov Chain in a way that it converges to the desired distribution, different probabilistic schemes have been developed. One of these methods is the Metropolis-Hasting MCMC method [115], which uses an accept-reject sampling scheme to control the sequence in a way that probability samples the target distribution after a certain point in the chain.

In this section we first review the concept of the Monte Carlo sampling and the

Markov chains. Then we explain the Metropolis-Hasting MCMC sampling algorithm and study different convergence tests to guarantee that a Markov chain has reached its equilibrium state. Our presentation of the MCMC methods here closely follows the exposition of Andrieu et al. [7].

4.3.1 Monte Carlo

The Monte Carlo approach was originally developed by physicist to generate random numbers for approximating complex integrals [7]. Suppose we want to estimate a complex integral

$$\int_a^b h(x)dx, \tag{4.28}$$

If we $h(x)$ can be decomposed into the production of a function $f(x)$ and a probability density function $p(x)$ defined over the interval (a, b), then we can write the equation (4.28)

$$\int_a^b h(x)dx = \int_a^b f(x)p(x)dx = E_{p(x)}[f(x)]. \tag{4.29}$$

Accordingly the integral can be expressed as an expectation of $f(x)$ over the probability density function $p(x)$. Thus, if we draw a large number of samples $x_1, x_2, \ldots, x_n$ from the probability density function $p(x)$, we can approximate the integral as

$$\int_a^b h(x)dx = E_{p(x)}[f(x)] \simeq \frac{1}{n}\sum_{i=1}^{n} f(x_i). \tag{4.30}$$

This is referred to as Monte Carlo integration.

4.3.2 Markov Chains

Let X_t denote the value of a random variable X at time t, and the state space refers to the range of possible values for X values. A random process is a Markov process if the transition probabilities between different values in the state space depend only on the random variable's current state, i.e.,

$$\Pr(X_{t+1} = s_j | X_0 = s_0, \ldots, X_t = s_i) = \Pr(X_{t+1} = s_j | X_t = s_i) \tag{4.31}$$

Therefore, a Markov process has no memory, in the sense that the only information required to predict the next state is the current state of the random variable. A Markov chain is then defined as a sequence of random variables $(X_0, \ldots, X_n)$ generated by a Markov process. Particularly, we define a Markov chain by its transition probability (or the transition kernel), which is the probability of a Markov process to have the value s_j, given the value of the previous step s_i

$$P(i,j) = P(i \to j) = \Pr(X_{t+1} = s_j | X_t = s_i). \tag{4.32}$$

We define $\pi_j(t)$ as the probability that the Markov chain has the value s_j at time t

$$\pi_j(t) = \Pr(X_t = s_j). \tag{4.33}$$

The probability of the Markov chain to have value s_i at time $t+1$ can be calculated with the Chapman-Kolmogorov equation [132] by summing over the conditional probabilities of all the possible moves from the previous state to have state value s_i

$$\begin{aligned} \pi_i(t+1) &= \Pr(X_{t+1} = s_i) \\ &= \sum_k \Pr(X_{t+1} = s_i | X_t = s_k).\Pr(X_t = s_k) \\ &= \sum_k P(k \to i)\pi_k(t) = \sum_k P(k,i)\pi_k(t). \end{aligned} \tag{4.34}$$

The equation (4.34) can be written in the matrix form as:

$$\boldsymbol{\pi}(t+1) = \boldsymbol{\pi}(t)\mathbf{P}. \tag{4.35}$$

If the start of the Markov chain is specified by a starting vector $\pi(0)$, the equation (4.35) can be recursively expressed as:

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(t-1)\mathbf{P} = (\boldsymbol{\pi}(t-2)\mathbf{P})\mathbf{P} = \boldsymbol{\pi}(t-2)\mathbf{P}^2 \ldots = \pi(0)\mathbf{P}^t. \tag{4.36}$$

Accordingly, we define the n-step transition probability $p_{ij}^{(n)}$ as the probability of a Markov process to have value s_j at time $t+n$ given the fact that the chain had value s_i at time t, i.e.,

$$p_{ij}^{(n)} = \Pr(X_{t+n} = s_j | X_t = s_i). \tag{4.37}$$

It can be easily seen the $p_{ij}^{(n)}$ is the $ij-$th element of $\mathbf{P}^n$

A Markov chain is **irreducible** if there exists a positive integer α such that $p_{ij}^{(\alpha)} > 0$ for all i, j. It guarantees that the chain can possibly move from any state to any other state. Similarly, a Markov chain is **aperiodic** if the chain is not forced into some cycle

of fixed length to move between any two states [7]. If a Markov chain is irreducible and aperiodic, it can be shown [7, 132] that a sufficient amount of time, it reaches a stationary distribution π^* that:

$$\pi^* = \pi^* \mathbf{P} \tag{4.38}$$

It implies that the probability of a Markov chain to be at any particular given state becomes independent of the initial condition when the chain reaches its stationary state.

The basic idea of discrete Markov chain can be generalized to a continuous Markov process that the transition kernel probability $P(x, y)$ satisfies

$$\int P(x, y)dy = 1, \tag{4.39}$$

and the continuous extension of the Chapman-Kolmogorov equation becomes

$$\pi_t(y) = \int \pi_{t-1}(x)P(x, y)dy. \tag{4.40}$$

Accordingly, when a continuous Markov chain reaches its stationary state, we have:

$$\pi^*(y) = \int \pi^*(x)P(x, y)dy. \tag{4.41}$$

4.3.3 The Metropolis-Hasting Sampling

The Metropolis (or Metropolis-Hastings) algorithm was developed by Metropolis and Ulam [116], Metropolis et al. (1953) [115], and Hastings (1970) [70]. It belongs to the family of MCMC methods, i.e., it is random (Monte Carlo) and also has no memory (Markov chain), in the sense that each step depends only on the previous step.

The basic idea is to modify a random walk, using a probabilistic rule, that some proposed moves are accepted and the others are rejected, to sample from a target distribution $p(\theta)$. For this purpose, an irreducible and aperiodic Markov chain is conducted in a way that at equilibrium state, it converges to the target distribution $p(\theta)$. The only requirement that the Metropolis-Hasting algorithm imposes is that a function proportional to $p(\theta)$ has to be calculable. In the Bayesian application, the normalization constant is often very difficult to compute. Therefore, the ability to draw samples from a not normalized distribution function is an important feature of the Metropolis-Hasting algorithm.

Assume we want to sample from $p(\theta)$ where $p(\theta) = \frac{f(\theta)}{K}$, where K is the normalizing constant and it is either not known or extremely difficult to compute. The Metropolis algorithm [116, 115] draws sample from $p(x)$ with the following procedure:

1. Start with any initial value θ_0, where $f(\theta_0) > 0$.

2. Using a proper proposal distribution $q(\theta_1, \theta_2)$ to generate a candidate sample θ^*. The only restriction on the proposal distribution in the Metropolis algorithm is that it has to be symmetric, i.e., $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.

3. Calculate the ratio of the probability of the new proposed sample to the probability of the current state:

$$\alpha = \frac{p(\theta^*)}{p(\theta_{t-1})} = \frac{f(\theta^*)}{f(\theta_{t-1})} \tag{4.42}$$

   Note that the normalization constant K cancels out in this fashion.

4. If the new proposed sample is more probable than the current sample ($\alpha > 1$), accept the move, otherwise with the probability α accept the proposed sample.

5. Repeat the procedure until the equilibrium state is achieved. We will discuss about the convergence diagnostic tools to check if a Markov chain has reached its equilibrium state in Section 4.3.4.

The above procedure can be summarized as first computing

$$\alpha = \min\Big(\frac{f(\theta^*)}{f(\theta_{t-1})}, 1\Big) \tag{4.43}$$

and then accepting a proposed move with probability α. This produces a Markov chain that the transition probability depends only on the current state and not the past states. It can be shown that after a sufficient "burn-in" steps, this Markov chain approaches its stationary distribution that is equal to $p(\theta)$.

Hastings (1970) generalized the Metropolis algorithm by assuming that the proposal distribution can be an arbitrary probability density function $q(\theta_1, \theta_2)$, and setting the acceptance probability of a proposed sample as:

$$\alpha = \min\Big(\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})q(\theta_{t-1}, \theta^*)}, 1\Big) \tag{4.44}$$

This is called the Metropolis-Hastings algorithm. If we assume that the proposal distribution is symmetric, i.e., $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$, the original Metropolis algorithm is then recovered. It can be shown coherently that the Metropolis-Hasting Markov chain converges to $p(\theta)$ at equilibrium [116, 115].

A key issue in the performance of the Metropolis-Hasting algorithm is the number of steps required for the Markov chain to converge to its stationary distribution (the burn-in period). Typically the first 1000 to 5000 samples are ignored and the convergence tests are used to examine whether the chain is converged [149]. A poor choice of starting

value or proposal distribution can considerably increase the burn-in time. The optimal choice for the starting point and the proposal distribution has been an area of current research [149]. There are two general approaches; random walks and independent chain sampling [7]. In the random walks approach the new proposed sample y equals to the current sample x plus a random variable z,

$$y = x + z. \tag{4.45}$$

Hence, $q(x, y) = h(y - x) = h(z)$, and any arbitrary probability distribution can be used for $h(z)$. In the independent chain approach, the probability of the new sample is independent of the current position x, i.e., $q(x, y) = h(y)$. One popular choice for h is the multivariate Gaussian distribution.

A chain can have a poor mixing, in the sense that it stays in small regions of the state space for a long period of time, as opposed to a well mixing chain that broadly probes the state space. The proposal distribution can be tuned to adjust the mixing of the chain and consequently the acceptance rate of the algorithm. This is generally done by adjusting the variance of the proposal distribution. However, there is a trade-off in choosing the variance of the proposal distribution. Draper [21] discussed that if the variance is too large, proposed moves are large and can broadly explore the space but not often accepted. On the other hand, if the variance of the proposal distribution is too small, moves are generally accepted but stays in a small region of the state space which results in poor mixing.

4.3.4 Convergence Diagnostics of Markov Chain

Diagnostic tools are usually used for evaluating the sampling performance of the chains drawn by the MCMC methods to determine whether the Markov chain has reached its stationary distribution. Consider a sequence of $(\theta_1, \ldots, \theta_n)$ of length n from a Metropolis-Hasting sequence. The adjacent samples in the chain are expected to be correlated. We quantify the correlation by using an autocorrelation function [34, 62]. The correlation between two samples in the chain $\rho(\theta_i, \theta_{i+k})$ can be estimated as

$$\hat{\rho_k} = \frac{\mathrm{Cov}(\theta_t, \theta_{t+k})}{\mathrm{Var}(\theta_t)} = \frac{\sum_{t=1}^{n-k}(\theta_t - \bar{\theta})(\theta_{t-k} - \bar{\theta})}{\sum_{t=1}^{n-k}(\theta_t - \bar{\theta})^2}, \quad \text{with} \quad \bar{\theta} = \frac{1}{n}\sum_{t=1}^{n}\theta_t. \tag{4.46}$$

This is the k^{th} order autocorrelation. An ideal sampler will have an autocorrelation that decays rapidly. High autocorrelations within the Markov chain illustrates poor mixing and subsequently slow convergence.

Another measure to test the convergence of the Markov chains is the power spectrum analysis [34]. The power spectrum of a finite MCMC chain is described by

$$P(k) = |a_k|^2, \tag{4.47}$$

in which a_k for $k = (\frac{2\pi j}{N})$ and $\{j = 1, \ldots, \lfloor \frac{N}{2} \rfloor - 1\}$ are the discrete inverse Fourier transform of the chain divided by the square root of the number of samples in the MCMC chain. An ideal sampler will have a flat power spectrum as the correlations within the Markov chain result in curvature on small scale levels of the power spectrum [43]. The statistical efficiency of an MCMC chain is commonly described as:

$$E = \frac{\sigma_T^2}{P_0}. \tag{4.48}$$

in which σ_T^2 is the variance of the target distribution and P_0 is the spectral density of the Markov at frequency zero $P(k = 0)$. E^{-1} heuristically determines the fraction that makes the Markov chain longer than an ideal chain. Several other methods have also been proposed in the literature for the convergence test of a Markov chain. A comparative review of these methods has been presented by Cowles et al. [34].

4.4 Case Study

In this section, we apply the Metropolis-Hasting MCMC algorithm using the polynomial chaos proxy described earlier for the water flooding of two synthetic channel sand models, similar to the synthetic model studied in Chapter 3.

In the first case study, the proposed procedure is tested in a simple synthetic reservoir problem. The second case study tends to be more challenging as the solution involves two different scenarios which makes the reservoir uncertainty quantification problem considerably nonlinear.

Case I

The simulation model represents a 2D horizontal square reservoir with two water injectors in the left corner and two producers in the right corner. The reservoir covers an area of $500 \times 500\text{m}^2$ and has a thickness of 10m. It is discretized into a 50×50 horizontal 2D grid. It is essentially an incompressible two-phase unit mobility oil-water system, with zero connate water saturation and zero oil residual system. The model is run for 1000 days with the injectors under rate control and the producers under bottom hole pressure (BHP) control with predefined rates and BHPs. The injection rate is 500 $\frac{m^3}{day}$ for each injectors. The objective is to obtain samples from the posterior distribution of the unknown permeability fields using observed data that entails the water cuts and oil flow rates for 1000 days.

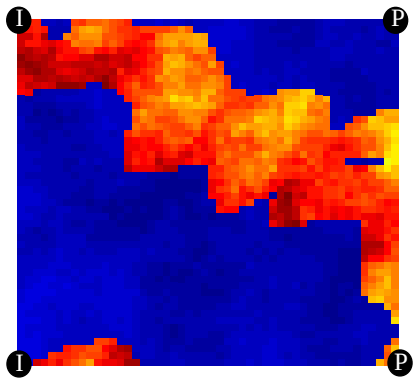

Figure 4.4: The synthetic water flooding model with two injectors on the left and two producers on the right. It also shows the reference permeability field used to generate the observed data of oil and water flow rates.

Although the permeability field is unknown, we assume that prior geological knowledge indicates that this is a fluvial channelized reservoir, as illustrated in the training

image in Figure 4.5, which has a very fine sand texture of permeability about 50mD, and the background with permeability about 10mD. The sand texture has a small variance of 5mD and the variance of the permeability of the background is 1mD. The porosity is assumed to be 0.3 all over the field. The problem is mainly to estimate the shape of the channel. Figure 4.5 shows the training image for the prior model of the fluvial channel.

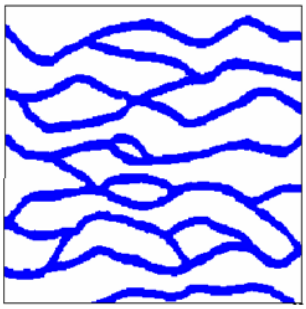

Figure 4.5: The training image for the fluvial channelized reservoir

Fig 4.6 illustrates the water flooding process of the fluvial channel for 1000 days. The observed data in Figure 4.7 includes the history of water and oil production of two producers for 1000 days. The true permeability model which is used as a reference case to generate the production data is shown in Figure 4.4.

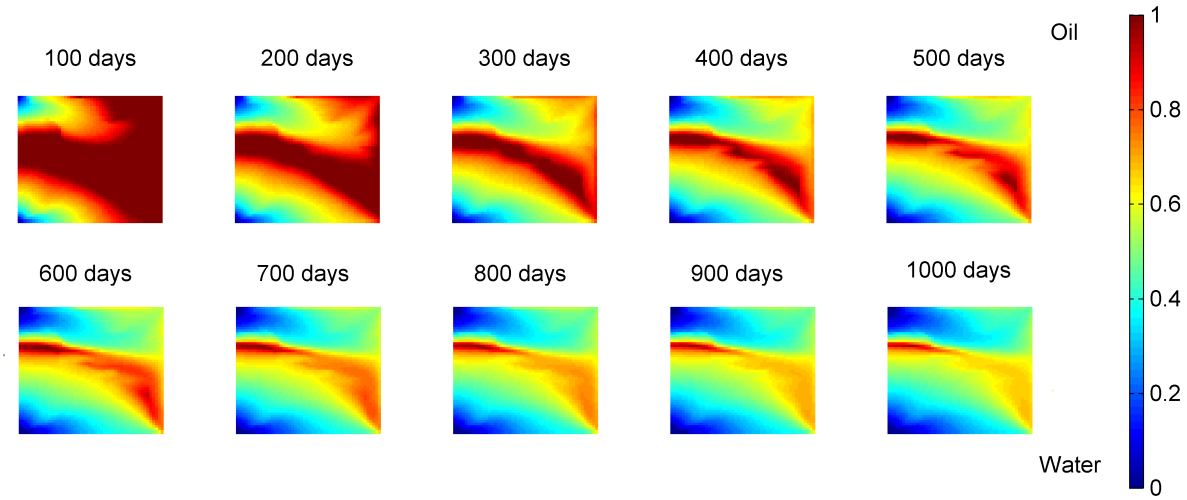

Figure 4.6: The water flooding process of the fluvial channel case I for 1000 days.

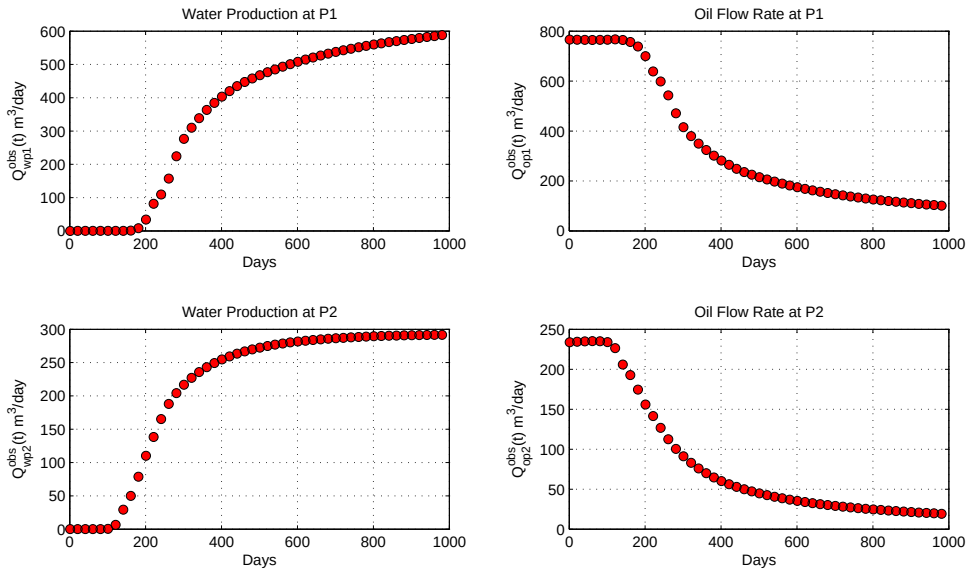

Figure 4.7: The observed history of the reservoir (before adding noise)

In Chapter 3 we the used polynomial chaos proxy to find the model which gives the global minimum to the misfit surface. Here we assume that the observations are not exact and include some measurement errors. Recalling the misfit formulation of (4.24), we assume the observational error $\sigma(t)$ for each oil and water flow rates $Q_{obs}(t)$ is

$$\sigma(t) = \sigma_D \, Q_{obs}(t) \tag{4.49}$$

where σ_D is constant. We solve the problem for $\sigma_D^2 = 0.1$ and $\sigma_D^2 = 0.2$. We also assume that the minimum noise (σ_{min}) is 30 $\frac{m^3}{day}$. In the Bayesian context, $\sigma^2(t)$ represents the standard deviation of the noise in data. Therefore it essentially has the same dimension of oil or water flow rates. Figure 4.8 illustrates the process of adding noise σ to data for the Bayesian inference method.

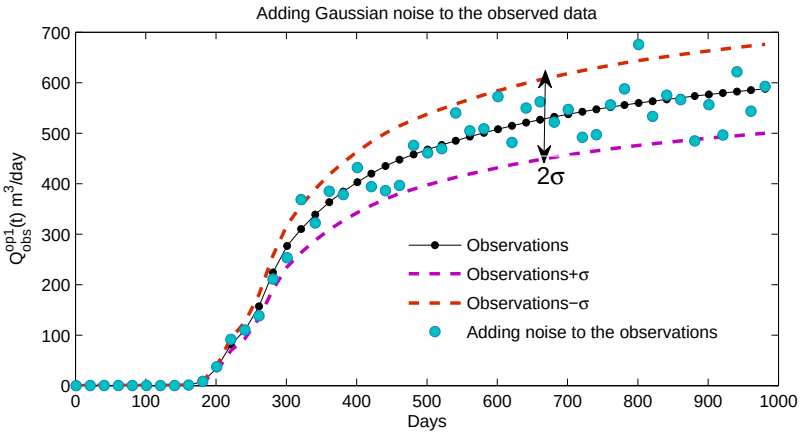

Figure 4.8: In Bayesian inference method we assume observations are not exact and include the observational errors $\sigma(t)$

To assess the uncertainty associated with the permeability field, using the polynomial chaos proxy, we have to follow three main steps. First, it is necessary to reduce the dimension of the problem to be able to efficiently construct the proxy model. Second we form the polynomial chaos proxy and calculate the posterior distribution with the Bayesian inference. The last step is to apply the Metropolis-Hasting MCMC to sample from the posterior distribution. The procedure of the polynomial chaos proxy-based uncertainty quantification is as follows:

- **Reduced-order parameterization of the permeability field**
  Using Petrel (Schlumberger) with the training image of Figure 4.5, we produce 1000 realizations, and compute the ensemble covariance matrix.

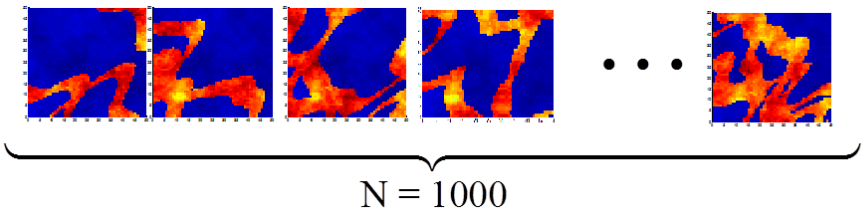

Figure 4.9: To train the polynomial chaos proxy, **N** realizations are generated based on the given prior distribution.

To reduce dimensions, we perform Karhunen-Loeve transform (linear PCA) using the ensemble covariance matrix and discard the smallest eigenvalues. Small eigenvalues incorporate high frequency changes in the permeability field. Figure 4.10 shows the eigenvalues of the ensemble covariance matrix of 1000 realizations. In

the statistics literature, this is also known as the "Scree graph" [85]. The optimum dimension of the system is often located after the "knee point" (or "elbow point"), where the Scree plot shows a significant drop in the eigenvalues [85]. The leading eigenvalue of the remaining ones is a good index of the error encountered in dimension reduction. Figure 4.10 illustrates a considerable drop in the eigenvalues after the 4^{th} eigenvalue. Hence, to preserve the shape of the channel, it suffices to keep the 4 largest eigenvalues and discard the rest. Since in our problem the transition between the two different textures happens sharply (high frequency change), discarding the smallest eigenvalues will result in having a continuous transition from sand to shale. As far as our problem concerns the general shape of the fluvial channel, we may neglect the high frequency changes and discard the small eigenvalues. We will also show that retaining 40 eigenvalues gives a slightly better illustration of sand and shale zones. Moreover, we apply the third order polynomial kernel PCA (see Chapter 3) with 4 eigenvectors retained in the feature space and compare the final results with linear PCA.

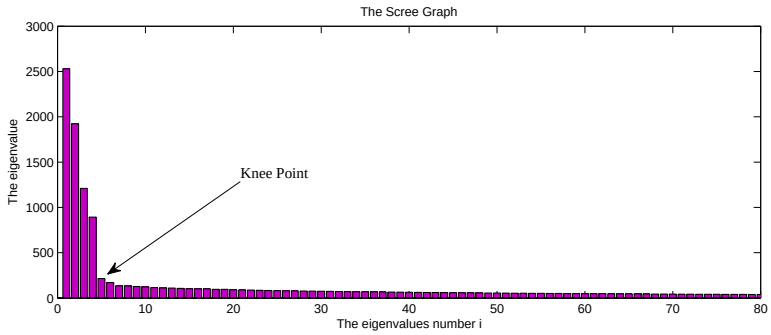

Figure 4.10: The Scree plot for the ensemble covariance matrix of 1000 realizations

- **Construct the polynomial chaos proxy**

  To train the polynomial chaos proxy, we run the reservoir simulator for each training realization of permeability, and obtain the flow rates of oil and water for both of the producers. For each realization $K_j(x, y)$ $(j = 1, \ldots, N)$ the reservoir simulator gives $\mathbf{Q^j_{op1}(t)}, \mathbf{Q^j_{wp1}(t)}, \mathbf{Q^j_{op2}(t)}, \mathbf{Q^j_{wp2}(t)}$. The polynomial chaos representation for oil and water flow rate is written as:

$$\begin{aligned}
\mathbf{Q^{op1}}(\xi,\mathbf{t}) &= Q_0^{op1} + \sum_{i=1}^{P} Q_{i,1}^{op1}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{op1}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{op2}}(\xi,\mathbf{t}) &= Q_0^{op2} + \sum_{i=1}^{P} Q_{i,1}^{op2}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op2}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{op2}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{wp1}}(\xi,\mathbf{t}) &= Q_0^{wp1} + \sum_{i=1}^{P} Q_{i,1}^{wp1}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{wp1}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{wp1}(\xi_i\xi_j) + \dots \\
\mathbf{Q^{wp2}}(\xi,\mathbf{t}) &= Q_0^{wp2} + \sum_{i=1}^{P} Q_{i,1}^{wp2}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{wp2}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{wp2}(\xi_i\xi_j) + \dots
\end{aligned} \tag{4.50}$$

The unknowns in the equation (4.50) are the polynomial chaos coefficients. Any of the numerical methods discussed in Section 2.4 can be applied to calculate the unknown coefficients of the equation (4.50) based on the $\mathbf{N}$ training data ($\mathbf{Q^j_{op1}}(\mathbf{t})$, $\mathbf{Q^j_{wp1}}(\mathbf{t})$, $\mathbf{Q^j_{op2}}(\mathbf{t})$, $\mathbf{Q^j_{wp2}}(\mathbf{t})$ for $j = 1, \dots, N$). We use regression-based PCM to calculate the polynomial chaos coefficients in a way that minimizes the sum of squared errors. The error is defined as the difference between the full reservoir simulator results and the fitted value provided by the polynomial chaos proxy $\mathbf{N} = 1000$ training data.

If we retain four largest eigenvalues $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, we will have four random variables $(\xi_1, \xi_2, \xi_3, \xi_4)$ in the polynomial chaos representation. Next, we have to decide on the order of the polynomial chaos expansion. The number of unknown coefficients in the polynomial chaos expansion of order p with M random variables is calculated as:

$$N_{PCE} = \sum_{K=0}^{p} \frac{(M+K)!}{M!K!}. \tag{4.51}$$

For example, the 8^{th} order polynomial chaos expansion with 4 random variables has 495 terms. Accordingly, the number of training simulation runs ($N = 1000$) is more than the number of unknown coefficients, if we use the 8^{th} order polynomial chaos expansion. Regression-based PCM is a standard approach to approximate solution of over-determined systems, i.e., set of equations in which there are more equations than unknowns. If we increase the order of the polynomial chaos expansion to 10, the number of unknown coefficients will be raised to 1001 and the system becomes slightly under-determined but can be readily solved by regression-based PCM with 1000 training runs to calculate the polynomial chaos coefficients.

Consequently, $N = 1000$ training simulation runs are quite sufficient to construct the higher order polynomial chaos proxy with 4 random variables. The number of term in the polynomial chaos expansion is inextricably linked with the number of eigenvectors retained and unfortunately it increases drastically with the increase of dimensionality. For example, if 40 eigenvectors are retained, the second order of the polynomial chaos expansion has 861 terms while the number of the unknown coefficients for the third order is raised to 12341. When the number of the training simulation runs is too small compared to the number of the terms (unknown coefficients), the regression problem in PCM is severely under-determined and non-unique. Accordingly, the number of the reservoir simulation runs required to compute the unknown coefficients of the polynomial chaos expansion increases greatly. As discussed in Chapter 2, several authors have proposed different methods to mitigate the "curse of dimensionality", i.e., Li et al. [100] preserved only the pure terms and discarded the cross terms. This reduces the number of coefficients remarkably and seems to give good results on some examples presented in their work. However, this will only work if the true forward model does not contain such nonlinear cross-effects of input random variables, which is not generally true for reservoir simulation models. Even more generally, using this approach invalidates one of the main advantages of the polynomial chaos proxy that guarantees convergence to the true distribution as the order of expansion increases. Subsequently, sparse polynomial chaos expansion was proposed [15, 16, 17] that is motivated by the so-called "sparsity of effects principle" [120]. The *sparsity of effects principle* states that the most models are mainly governed by main-effects and only the low-order interactions between input random variables [120]. Their proposed algorithm to achieve the sparse polynomial chaos expansion is to select the high-order polynomials in one single random variable and then use the cross-validation technique to determine the low-order interaction terms that should be preserved [15].

The approximation we propose here is the **reduced-terms polynomial chaos proxy**, which uses the impact factor to discard the interaction terms that are not relevant. We use the facts that the random variables ξ_i in the polynomial chaos expansion when applied along with PCA, are not of the same priority to uncertainty quantification problem as they are weighted by $\sqrt{\lambda_i}$ in the reduced-dimension representation of the random field. Considering the Karhunen-Loeve

representation of the permeability field [75]

$$\mathbf{K} = \bar{\mathbf{K}} + \sum_{i=1}^{p} \sqrt{\lambda_i} f_i \xi_i, \tag{4.52}$$

each random variable is multiplied by the factor $\sqrt{\lambda_i} f_i$ while $\|f_i\|_2 = 1$. Hence, any functional of $\mathbf{g}(\mathbf{K})$ is a function of the linear combinations of random variables ξ_i with the aforementioned weighting. Consequently, in the n^{th} order polynomial chaos representation of $\mathbf{g}(\mathbf{K})$, we expect that the coefficient associated with the term ξ_1^n to the coefficient associated with the term ξ_2^n to be intuitively attenuated by the factor $\left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}\right)^n$. To demonstrate this, the coefficients of the term ξ_1^n in the polynomial expansion [64] of $\mathbf{g}(\mathbf{K}(\xi))$ can be calculated as:

$$\begin{aligned} a_{\xi_1^n} &= \frac{1}{n!}\Big(\frac{\partial^n g}{\partial \xi_1^n}\Big)\Big|_{\xi_1=0} = \frac{1}{n!}\Big(\frac{\partial^n g}{\partial \mathbf{K^n}}\frac{\partial^n \mathbf{K}}{\partial \xi_1^n}\Big)\Big|_{\xi_1=0} \\ &= (\sqrt{\lambda_1})^n \frac{1}{n!}\Big(\frac{\partial^n g}{\partial \overline{\mathbf{K}}^{\mathbf{n}}} \circ \underbrace{f_1 \circ f_1 \circ \ldots \circ f_1}_{n}\Big), \end{aligned} \tag{4.53}$$

where $f_1 \circ f_1 \circ \ldots \circ f_1$ is the Hadamard product or element-wise product, and $\frac{\partial^n g}{\partial \overline{\mathbf{K}}^{\mathbf{n}}}$ is the vector derivative taken with respect to the mean of the permeability field. The same could be achieved for $a_{\xi_2^n}$:

$$a_{\xi_2^n} = (\sqrt{\lambda_2})^n \frac{1}{n!}\Big(\frac{\partial^n g}{\partial \overline{\mathbf{K}}^{\mathbf{n}}} \circ \underbrace{f_2 \circ f_2 \circ \ldots \circ f_2}_{n}\Big). \tag{4.54}$$

Noting the fact that all entries of the eigenvectors f_i are smaller than 1 and also $\|f_i\| = 1$, the main difference between the equation (4.53) and (4.54) appears in the factor $\left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}\right)^n$. Since ξ_1^n and ξ_2^n are identically distributed, when $\left(\frac{\sqrt{\lambda_2}}{\sqrt{\lambda_1}}\right)^n \ll 1$ then $a_{\xi_2^n} \ll a_{\xi_1^n}$ and the term ξ_2^n could be neglected in approximation. Therefore, it is reasonable to drop high-order terms proportional to small eigenvalues in favor of the terms associated with large eigenvalues. Accordingly, the optimal reduced-terms polynomial chaos representation includes high-order terms of random variables weighted by large eigenvalues and only low-order terms of random variables weighted by small eigenvalues. The same approximation can be applied for the cross-terms. For example, the coefficients associated with the term $\xi_1^5 \xi_2^2$ can be accordingly derived as:

$$a_{\xi_1^5\xi_2^2} = \frac{1}{5!2!}\Big(\frac{\partial^7 g}{\partial \xi_1^5 \partial \xi_2^2}\Big)\Big|_{\xi=0} = (\sqrt{\lambda_1})^5(\sqrt{\lambda_2})^2 \frac{1}{5!2!}\Big(\frac{\partial^7 g}{\partial \overline{\mathbf{K}}^{\mathbf{7}}} \circ \underbrace{f_1 \circ \ldots \circ f_1}_{5} \circ f_2 \circ f_2\Big). \tag{4.55}$$

While the terms $\xi_1^5\xi_2^2$ and $\xi_1^2\xi_2^5$ are identically distributed, the coefficient of the term $\xi_1^5\xi_2^2$ is approximately $\frac{(\sqrt{\lambda_1})^5(\sqrt{\lambda_2})^2}{(\sqrt{\lambda_1})^2(\sqrt{\lambda_2})^5} = \frac{(\sqrt{\lambda_1})^3}{(\sqrt{\lambda_2})^3}$ times greater than the coefficient of the term $\xi_1^2\xi_2^5$. Hence, we can drop the term $\xi_1^2\xi_2^5$ and preserves the term $\xi_1^5\xi_2^2$ when $\frac{(\sqrt{\lambda_2})^3}{(\sqrt{\lambda_1})^3} \ll 1$. For all other cross-terms the same analogy could be applied. For each cross-term we compare it to the equally-distributed term with the largest coefficient and discard the negligible terms. For example, for the term $\xi_1\xi_2^4\xi_3^3\xi_4^2\xi_5$, the equally-distributed term with the maximum coefficient is $\xi_1^4\xi_2^3\xi_3^2\xi_4\xi_5$ whose corresponding coefficient is approximately $\frac{(\sqrt{\lambda_1})^4(\sqrt{\lambda_2})^3(\sqrt{\lambda_3})^2(\sqrt{\lambda_4})(\sqrt{\lambda_5})}{(\sqrt{\lambda_1})(\sqrt{\lambda_2})^4(\sqrt{\lambda_3})^3(\sqrt{\lambda_4})^2(\sqrt{\lambda_5})}$ times greater than the coefficient of the term $\xi_1\xi_2^4\xi_3^3\xi_4^2\xi_5$. If $\frac{(\sqrt{\lambda_2})(\sqrt{\lambda_3})(\sqrt{\lambda_4})}{(\sqrt{\lambda_1})^3} \ll 1$ we discard the term $\xi_1\xi_2^4\xi_3^3\xi_4^2\xi_5$ comparing to the equally-distributed term $\xi_1^4\xi_2^3\xi_3^2\xi_4\xi_5$.

To specify the selection criteria, we define the heuristic impact factor of the order m for the term $\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_p^{k_p}$ where $k_1 + k_2 + \ldots + k_p = m$ as:

$$\mathbf{I}_m(\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_p^{k_p}) = \frac{(\sqrt{\lambda_1})^{k_1}(\sqrt{\lambda_2})^{k_2}\ldots(\sqrt{\lambda_p})^{k_p}}{(\sqrt{\lambda_1})^{k_1^*}(\sqrt{\lambda_2})^{k_2^*}\ldots(\sqrt{\lambda_p})^{k_p^*}}, \tag{4.56}$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_p$ and $\{k_i^*\}$ is $\{k_i\}$ sorted in descending order. In the reduced-terms polynomial chaos representation, if the impact factor of the term $\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_p^{k_p}$ is smaller than the predetermined cut-off value ϵ, $\mathbf{I}_m(\xi_1^{k_1}\xi_2^{k_2}\ldots\xi_p^{k_p}) < \epsilon$, the term is discarded, otherwise it will be preserved in the polynomial chaos expansion. Accordingly, the impact factor of the term ξ_i^m is defined as:

$$\mathbf{I}_m(\xi_i^m) = \frac{(\sqrt{\lambda_i})^m}{(\sqrt{\lambda_1})^m}, \tag{4.57}$$

where λ_1 is the largest eigenvalue. As a result, the maximum order m of pure terms ξ_i^m in the reduced-terms polynomial chaos expansion is the maximum m while $\mathbf{I}_m(\xi_1^m) > \epsilon$. Here we make another approximation based on the law of "sparsity of effects principle" [120], that if the maximum order of ξ_i in the pure term is m, we neglect the higher order of ξ_i in the cross-terms. Additionally, only few of the remaining cross-terms are selected based on their impact factor in the equation (4.56). To illustrate the impact factor, let us interrupt the case study with a simple example of the reduced-terms polynomial chaos representation of a function of two Gaussian random variables.

Example 4.4.1. ***Reduced-terms polynomial chaos representation:***
Let x, y be two Gaussian random variables with zero mean and the covariance matrix of $\left(\begin{smallmatrix}1 & 0.8\\ 0.8 & 1\end{smallmatrix}\right)$. The eigenvalues of the covariance matrix are $\lambda_1 = 1.8$ and $\lambda_2 = 0.2$ and the eigenvectors are $f_1 = \left(\begin{smallmatrix}0.7071\\ 0.7071\end{smallmatrix}\right)$ and $f_2 = \left(\begin{smallmatrix}-0.7071\\ 0.7071\end{smallmatrix}\right)$. Let $f(x, y) =$

e^{-x-2y} be a function of x and y that we want to represent it with the reduced-terms polynomial chaos expansion. First, we write the Karhunen-Loeve representation of (x, y) as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \sqrt{1.8} \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix} \xi_1 + \sqrt{0.2} \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix} \xi_2 = \begin{bmatrix} 0.9487 \\ 0.9487 \end{bmatrix} \xi_1 + \begin{bmatrix} -0.3162 \\ 0.3162 \end{bmatrix} \xi_2. \tag{4.58}$$

Then $f(x, y) = e^{-x-2y}$ can be written as

$$\begin{aligned} f(x, y) = e^{-x-2y} &= e^{-(0.9487\xi_1 - 0.3162\xi_2) - 2(0.9481\xi_1 + 0.3162\xi_2)} \\ &= e^{2.8443\xi_1 - 0.3162\xi_2}. \end{aligned} \tag{4.59}$$

We use numerical integration of the equation (2.19) to compute the polynomial chaos coefficients. Considering the 4^{th} order terms in the polynomial chaos expansion:

$$\begin{aligned} PC_4(\xi_1, \xi_2) = {} & 163.7313(\xi_1^4 - 6\xi_1^2 + 3) + 0.025(\xi_2^4 - 6\xi_2^2 + 3) \\ & - 72.8078(\xi_1^3 - 3\xi_1)\xi_2 - 0.89899\xi_1(\xi_2^3 - 3\xi_2) \\ & + 12.1410(\xi_1^2 - 1)(\xi_2^2 - 1). \end{aligned} \tag{4.60}$$

ξ_1 and ξ_2 are independent random variables with zero mean and unit variance (energy). The polynomial chaos coefficients associated with $(\xi_1^4 - 6\xi_1^2 + 3)$ is $\frac{163.7313}{0.025} = 81\left(\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}\right)^4$ times greater than the polynomial chaos coefficient associated with $(\xi_2^4 - 6\xi_2^2 + 3)$. Hence, in the 4^{th} order terms of the polynomial chaos expansion, it is reasonable to discard the latter comparing to the former, considering that ξ_1 and ξ_2 are identically distributed. To decide on the maximum order of ξ_2 in the reduced-terms polynomial chaos expansion, we calculate the impact factor of ξ_2^m and find the maximum m that $\mathbf{I}_m(\xi_1^m) > \epsilon$:

$$\begin{aligned} \mathbf{I}_2(\xi_2^2) &= \frac{(\sqrt{\lambda_2})^2}{(\sqrt{\lambda_1})^2} = \frac{1}{9} \\ \mathbf{I}_3(\xi_2^3) &= \frac{(\sqrt{\lambda_2})^3}{(\sqrt{\lambda_1})^3} = \frac{1}{27} \\ \mathbf{I}_4(\xi_2^4) &= \frac{(\sqrt{\lambda_2})^4}{(\sqrt{\lambda_1})^4} = \frac{1}{81} \end{aligned} \tag{4.61}$$

If we choose the cut-off value of $\epsilon = 0.1$, the maximum order of ξ_2 in the reduced-terms polynomial chaos expansion will be 2. Then the 4^{th} order of the reduced-terms polynomial chaos representation will be

$$RPC_4(\xi_1, \xi_2) = 163.7313(\xi_1^4 - 6\xi_1^2 + 3) - 72.8078(\xi_1^3 - 3\xi_1)\xi_2 + + 12.1410(\xi_1^2 - 1)(\xi_2^2 - 1). \tag{4.62}$$

The impact factor heuristically specifies the terms that are more relevant to the polynomial chaos representation. Obviously, discarding all the cross terms as Li et al. [100] proposed, ignores the significant term $-72.8078(\xi_1^3 - 3\xi_1)\xi_2$ *and will lead to a considerable error in estimation of* $f(x, y)$*. The method we proposed here, intuitively preserves the cross-terms that are relevant.*

The idea of the reduced-terms polynomial chaos representation is basically the generalization of "Scree graph" in high order terms. In the sense that similar to Scree plot where we ignore the random variables (terms) weighted by the small eigenvalues, in higher order we discard the terms associated with powers of the small eigenvalues.

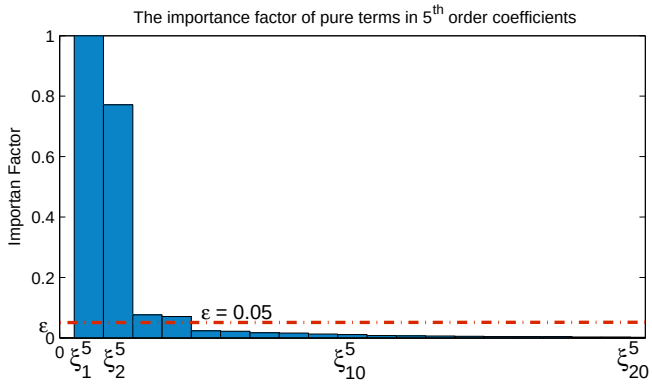

Figure 4.11: The impact factor of the 5^{th} order terms in the reduced-terms polynomial chaos expansion for the case study I

Returning to our case study, if 40 eigenvalues are retained, the polynomial chaos proxy of order 8 will have more than 30 million coefficients! It is obviously not feasible to run millions of reservoir simulations to calculate these coefficients. The solution lies in the fact that most of these coefficients are very small and could be discarded by calculating the impact factor. To illustrate that, consider Figure 4.11 that shows the impact factor of the 5^{th} order terms. The cut off value of $\epsilon = 0.05$ shows that only $\xi_1^5, \xi_2^5, \xi_3^5, \xi_4^5$ presents in the reduced-terms polynomial chaos expansion. Figure 4.12 shows the importance factor of 8^{th} order and with $\epsilon = 0.05$ only ξ_1^8, ξ_2^8 remains in the reduced-term representation. The cross-terms are not shown in Figure 4.11 and Figure 4.12 but the same routine also applies to them and we only select terms that their corresponding impact factor is higher than 0.05.

Consequently, if 40 eigenvalues are retained, the 8^{th} order reduced-terms polynomial chaos expansion can be approximately achieved by discarding the small terms with $N = 1000$ reservoir simulation runs. If we set $\epsilon = 0.05$, from more than 30 million terms, the impact factor only selects 1282 terms that are more relevant. Regression-based PCM can be used then to calculate the reduced-terms polynomial chaos expansion with $N = 1000$ reservoir simulation runs. In the reduced-terms representation, the maximum order of ξ_1 and ξ_2 is 8, the maximum order of ξ_3, ξ_4 is 6, the maximum order of $\xi_5, \ldots, \xi_{15}$ is 3, the maximum order of $\xi_{16}, \ldots, \xi_{30}$ is 2 and only the first order of all other random variables $\xi_{31}, \ldots, \xi_{40}$ is present. Hence, the random variables appear with different order in the reduced-terms polynomial chaos expansion. Similarly, only few dominant cross-terms are preserved in the reduced-term representation. As a result, the curse of dimensionality is mitigated to a great extent in the reduced-terms polynomial chaos expansion when applied along with the Karhunen-Loeve expansion (PCA).

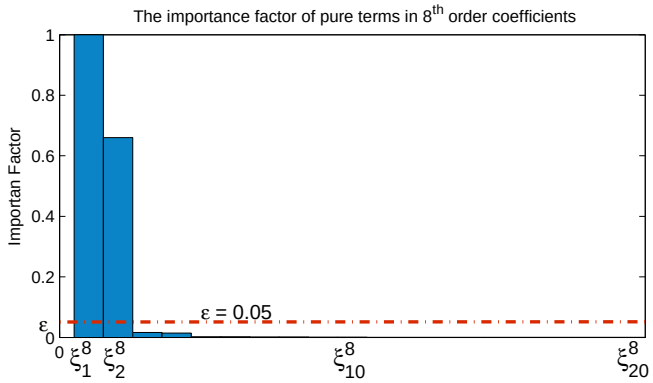

Figure 4.12: The impact factor of 8^{th} order terms in the reduced-terms polynomial chaos expansion for the case study I

Retaining 40 eigenvalues in PCA, clearly illustrates the shape of the fluvial channel in this case study and we do not need to use more than 40 random variables. However, if one decides to increase the resolution and retain more eigenvalues, only the first order of their corresponding random variables will appear in the reduced-terms polynomial chaos expansion, because their impact factor is smaller than $\mathbf{I}_m(\xi_{40}^m)$. Hence the number of coefficients in the reduced-terms polynomial chaos expansion increases linearly with the number of new added random variables.

The same procedure can be applied to construct the reduced-terms polynomial chaos expansion of random fields that are represented by kernel PCA. Because the random variables in the feature space is also weighted by $\sqrt{\lambda_i} f_i$, the terms in the

polynomial chaos expansion appears with different weighting. Hence the impact factor can be calculated to preserve only the relevant terms in the polynomial chaos representation of the output random variable.

- **Bayesian Inference Formulation**
  Using the misfit definition of the equation (4.25), as the difference between the observed and modeled production data, we have:

$$\begin{aligned}\mathbf{S}(\xi_{\mathbf{1}},\xi_{\mathbf{2}},\ldots,\xi_{\mathbf{p}}) = &\sum_{i=1}^{M}\frac{(\mathbf{Q}^{\mathbf{op1}}_{\mathbf{proxy}}(\mathbf{t_i})-\mathbf{Q}^{\mathbf{op1}}_{\mathbf{obs}}(\mathbf{t_i}))^2}{\sigma^2(t_i)}+\sum_{i=1}^{M}\frac{(\mathbf{Q}^{\mathbf{op2}}_{\mathbf{proxy}}(\mathbf{t_i})-\mathbf{Q}^{\mathbf{op2}}_{\mathbf{obs_i}}(\mathbf{t_i}))^2}{\sigma^2(t_i)}+\ldots\\ &+\sum_{i=1}^{M}\frac{(\mathbf{Q}^{\mathbf{wp1}}_{\mathbf{proxy}}(\mathbf{t_i})-\mathbf{Q}^{\mathbf{wp1}}_{\mathbf{obs}}(\mathbf{t_i}))^2}{\sigma^2(t_i)}+\sum_{i=1}^{M}\frac{(\mathbf{Q}^{\mathbf{wp2}}_{\mathbf{proxy}}(\mathbf{t_i})-\mathbf{Q}^{\mathbf{wp2}}_{\mathbf{obs}}(\mathbf{t_i}))^2}{\sigma^2(t_i)}.\end{aligned} \tag{4.63}$$

  If we plug the polynomial chaos expression of the equation (4.50) for water and oil flow rates of each producer, we obtain the following approximation for the misfit surface, in which the misfit coefficients $(S_0, S_{i,1}, S_{ii,2}, \ldots)$ are already known:

$$\mathbf{S}(\xi_{\mathbf{1}},\xi_{\mathbf{2}},\ldots,\xi_{\mathbf{p}}) = S_0+\sum_{i=1}^{p}S_{i,1}\xi_i+\sum_{i=1}^{p}S_{ii,2}(\xi_i^2-1)+\sum_{i=1}^{p-1}\sum_{j>i}^{p}S_{ij,2}(\xi_i\xi_j)+\ldots. \tag{4.64}$$

  The equation (4.65) gives an analytical approximation for the posterior probability distribution:

$$\sigma_\xi(\xi) = k\,\rho_\xi(\xi)\,\mathbf{exp}\{-(S_0+\sum_{i=1}^{p}S_{i,1}\xi_i+\sum_{i=1}^{p}S_{ii,2}(\xi_i^2-1)+\sum_{i=1}^{p-1}\sum_{j>i}^{p}S_{ij,2}\xi_i\xi_j+\ldots)\} \tag{4.65}$$

  If we assume the priors for ξ_i are all standard Gaussian;

$$\rho_\xi(\xi) = \Big(\frac{1}{\sqrt{2\pi}}\Big)^p e^{-\sum_{i=1}^{p}\frac{\xi_i^2}{2}}. \tag{4.66}$$

  We can sample from the posterior distribution of ξ using the Metropolis-Hasting MCMC. As we discussed earlier, in the Bayesian applications, the normalization factor is often very difficult to compute, and for that reason the ability to generate samples without knowing this constant of proportionality is an important advantage of the Metropolis-Hasting MCMC method.

- **Metropolis-Hasting MCMC**

  The final step is to sample from the posterior probability distribution of the reduced dimension parameters. We apply the Metropolis-Hasting MCMC method with a multivariate Gaussian distribution as the proposal distribution, and let the random walk reach its equilibrium state. To check if the chain has converged to the target distribution, we can use the methods introduced in Section 4.3.4. Alternatively, we can examine the posterior moments to check if all are converged. Figure 4.13 shows that the first two moments of the reduced-dimension parameters have converged after ten thousand runs of the algorithm. All other moments up to 8^{th} order are converged respectively and we are safe to say we have reached the equilibrium state for our desired precision. Although the higher order terms are not zero in the misfit formulation of (4.64), in the posterior formula (4.65) they are damped by the exponential function and becomes negligible. Hence, the posterior probability distribution of ξ is approximately a multivariate Gaussian with the posterior mean and covariance.

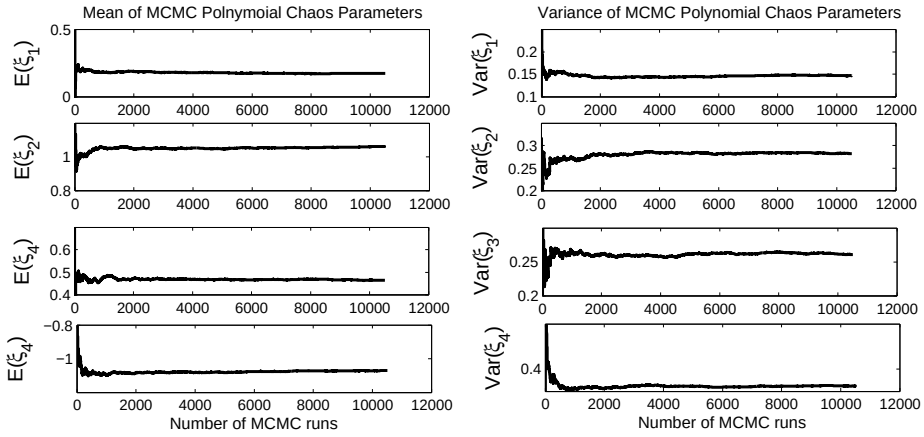

Figure 4.13: Convergence check for Metropolis-Hasting MCMC

Figure 4.14 shows the posterior probability distribution for the four dominant random variables of the feature space, compared to the true results obtained by the traditional MCMC method. Note that the correlations of the parameters are negligible in this problem. we allow the dimensionless error in the observation to be $\sigma_D^2 = 0.1$ in the equation (4.49). This has a direct effect on the posterior distribution function. In the sense that if we increase the variance of error in the observed data, the posterior variances of the reduced-dimension parameters increase. Figure 4.16 shows the posterior

distribution of the reduced-dimension parameters when the dimensionless variance of error in the observation is $\sigma_D^2 = 0.2$.

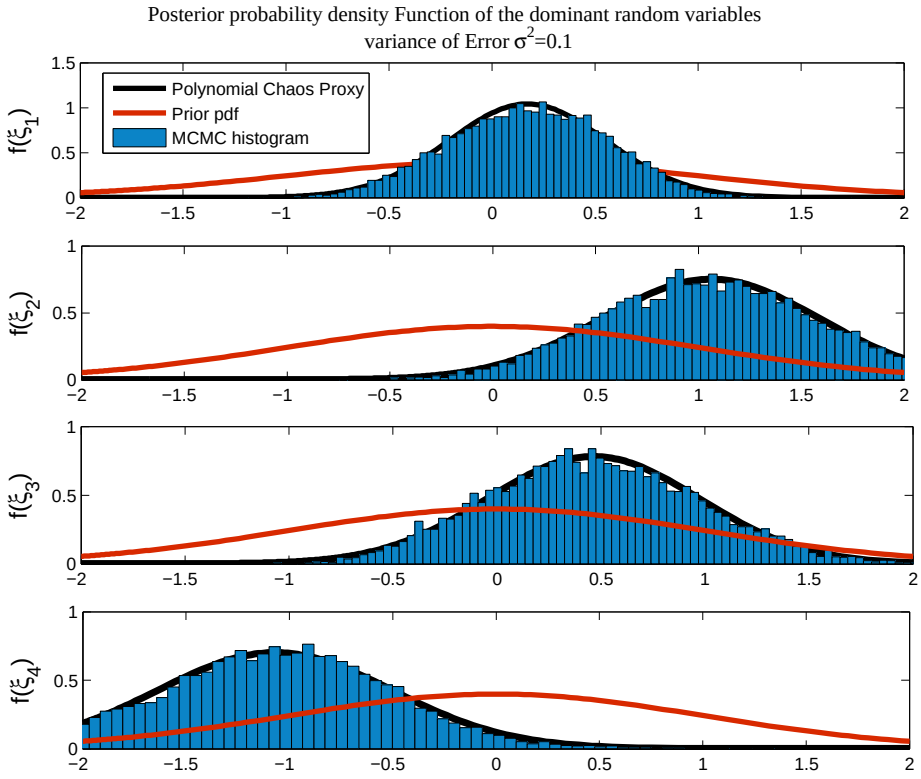

Figure 4.14: The posterior probability distribution for the reduced-dimension parameters when the dimensionless error in the observation is $\sigma_D^2 = 0.1$

To calculate the posterior distribution of the random variables associated with the 40 largest eigenvectors, more than 10^7 full reservoir simulation is required, which is not computationally practical. Hence we compare the accuracy of the polynomial chaos proxy in the approximation of the first two moments of the random variables with the Monte Carlo method. Figure 4.15 illustrates the absolute error in the approximation of the first two moments. It demonstrates that the accuracy decrease for the higher frequency modes, because in the reduced terms polynomial chaos expansion we have neglected some of the terms associated with them.

If the measurement error in the observation is narrowed down, the variance of the posterior distribution will decrease. In the simulated annealing technique, the variance of error in the observation, σ^2, is gradually reduced to zero, to find the global minimum of misfit surface [176]. If we allow no error in the observation, obviously the posterior distribution converges to a discrete point.

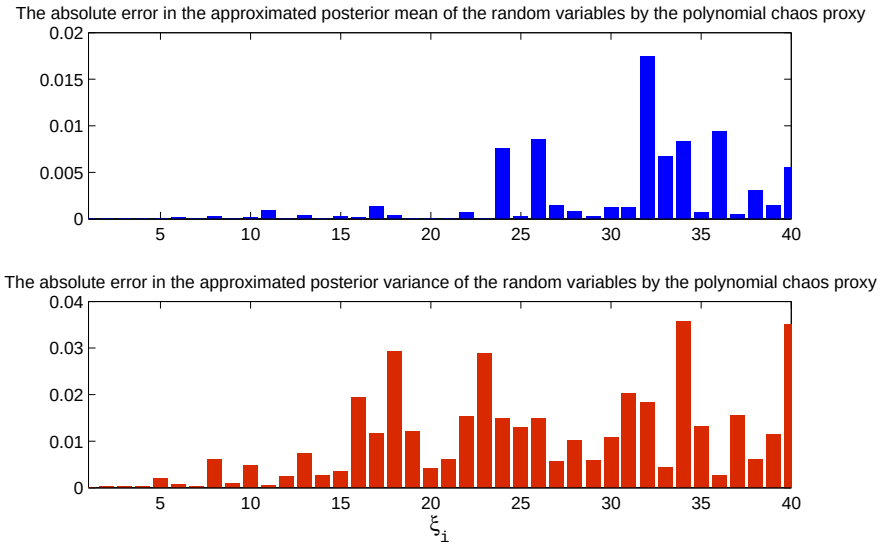

Figure 4.15: The absolute error in the polynomial chaos approximation of the first two moments compared to the Monte Carlo results when $\sigma_D^2 = 0.1$.

Since there is a map from the feature space of the reduced-order parameters to the high-resolution permeability field, we can obtain samples from the posterior distribution of the permeability field. By solving the pre-image problem with an iterative scheme, for each sample of the random variables in the feature space ξ, we generate samples from the posterior distribution of the permeability field $K(x, y)$. Using simple statistics, we can compute the posterior mean, covariance, etc., from the samples of the posterior distribution of $K(x, y)$. Figure 4.17 demonstrates the posterior mean of the permeability field, while 4 largest eigenvalues are retained.

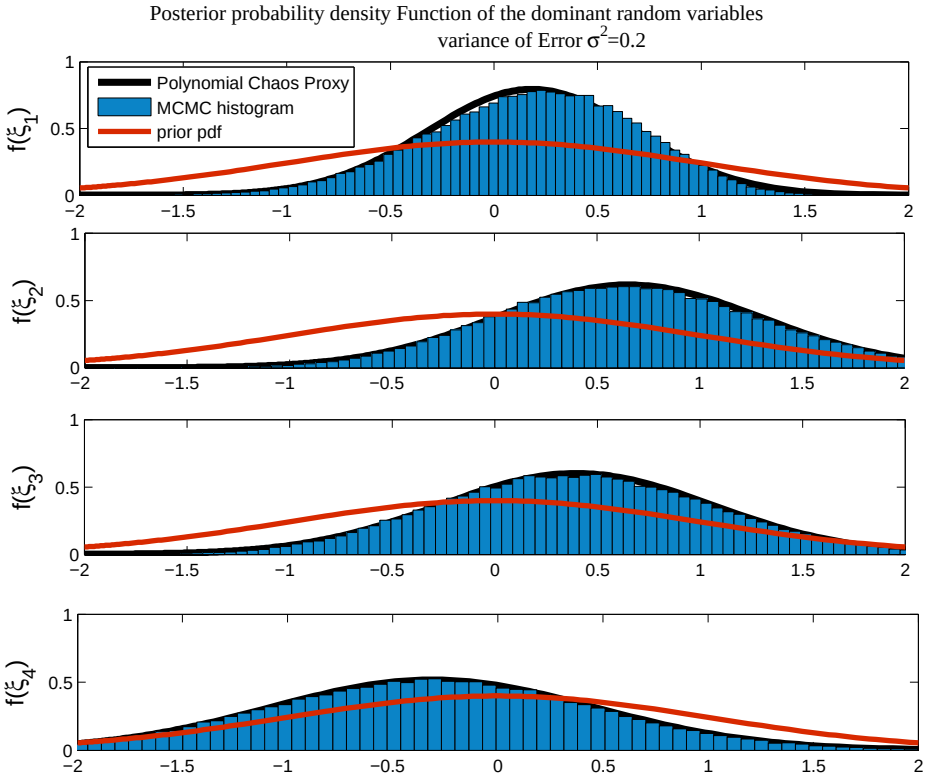

Figure 4.16: The posterior probability distribution for the reduced-dimension parameters when the dimensionless error in the observation is $\sigma_D^2 = 0.2$

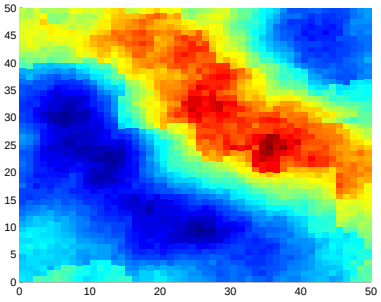

Figure 4.17: The posterior mean of the permeability field using 4 random variables with PCA

Figure 4.18 illustrates how the high-frequency change in the transition zone between sand and shale is preserved by retaining more eigenvalues. We construct the reduced-terms polynomial chaos proxy of 8^{th} order, with 5, 10, 30 and 50 random variables (which is equal to the number of eigenvectors retained). It also shows a comparison between

Karhunen-Loeve transformation (linear PCA) and the polynomial kernel PCA of order 3 results. Figure 4.18 indicates that the result of PCA and kernel PCA (3^{rd} order polynomial kernel) are not significantly different. It shows that the given realizations of the permeability field can seemingly be approximated with linear parametrization.

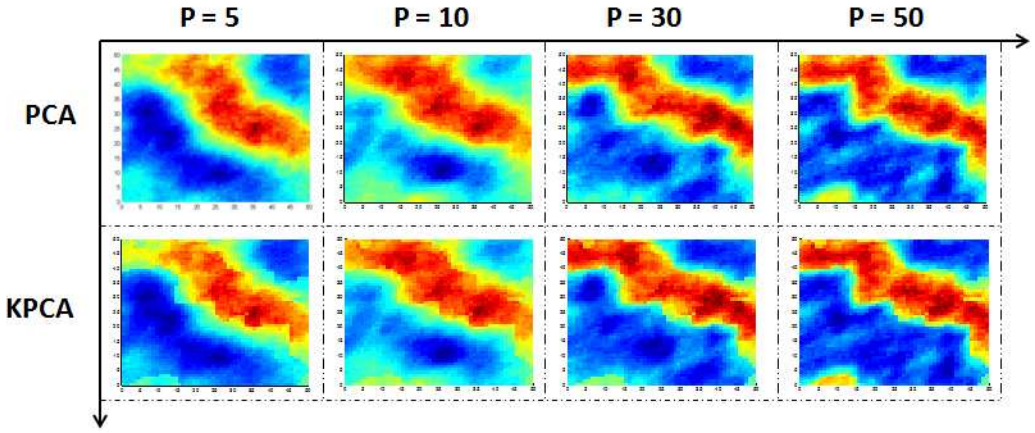

Figure 4.18: The posterior mean of the permeability field using PCA and kernel PCA decomposition method, while different number of eigenvectors (p=5, 10, 30, 50) in feature space is retained. The reduced-terms polynomial chaos proxy of order 8 is constructed for each case.

As far as the posterior probability distribution is concerned, the higher order terms in the polynomial chaos expansion in the misfit surface is attenuated by damping exponential function, $e^{-\mathbf{S}(\xi)}$. Hence, employing low orders of the polynomial chaos expansion usually gives a reasonable accuracy. Figure 4.19 and Figure 4.20 shows the full-simulator run for the samples obtained from the posterior distribution of the permeability field, when the dimensionless error in the observation is $\sigma_D^2 = 0.1$ and $\sigma_D^2 = 0.2$. Figures 4.21, 4.22, 4.23 and 4.24 show different samples from the posterior probability distribution of permeability field by the Metropolis-Hasting MCMC, when the dimensionless error in the observation of error is $\sigma_D^2 = 0.1$ and 0.2, and different number of eigenvectors is retained. For these figures we used the reduce-terms polynomial chaos proxy of order 8 and linear PCA is applied to reduce the dimension of the permeability field. Comparing the results of these figures, it shows that the number of eigenvectors retained has a considerable influence on the sampling resolution. It indicates that retaining four eigenvectors only preserves the general shape of the fluvial channel and it is not enough for high-resolution sampling. To retrieve fine-scale geology, we increased the number of eigenvectors to 40 in Figure 4.22 and Figure 4.24.

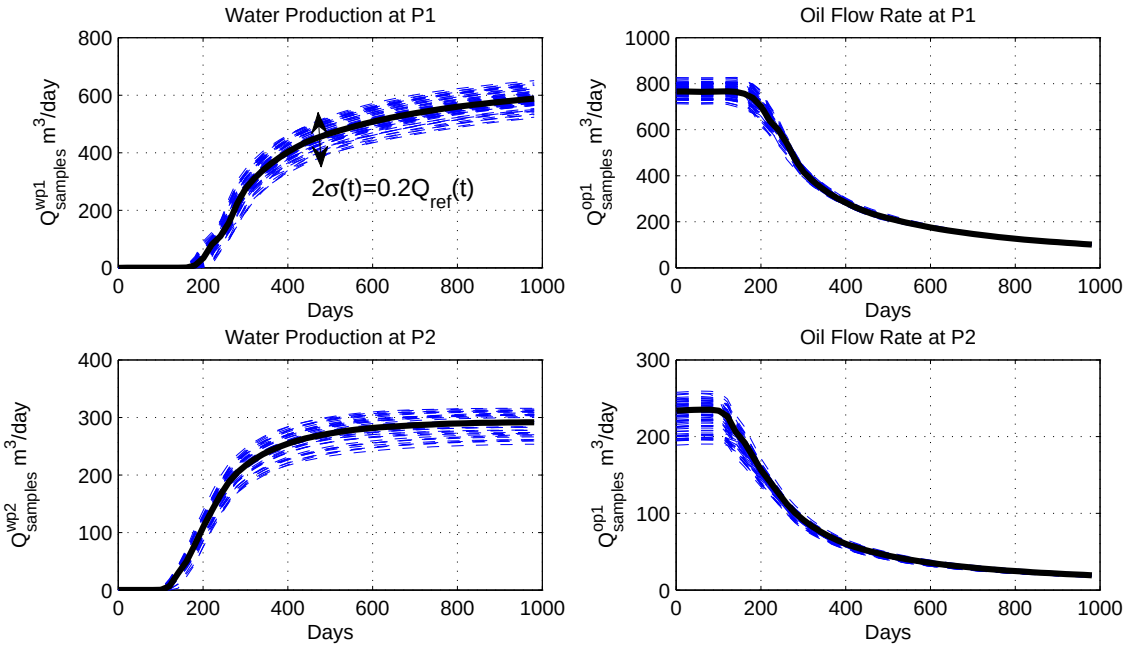

Figure 4.19: Full reservoir simulator runs for the samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.1$, vs. the reference case history (solid black curves).

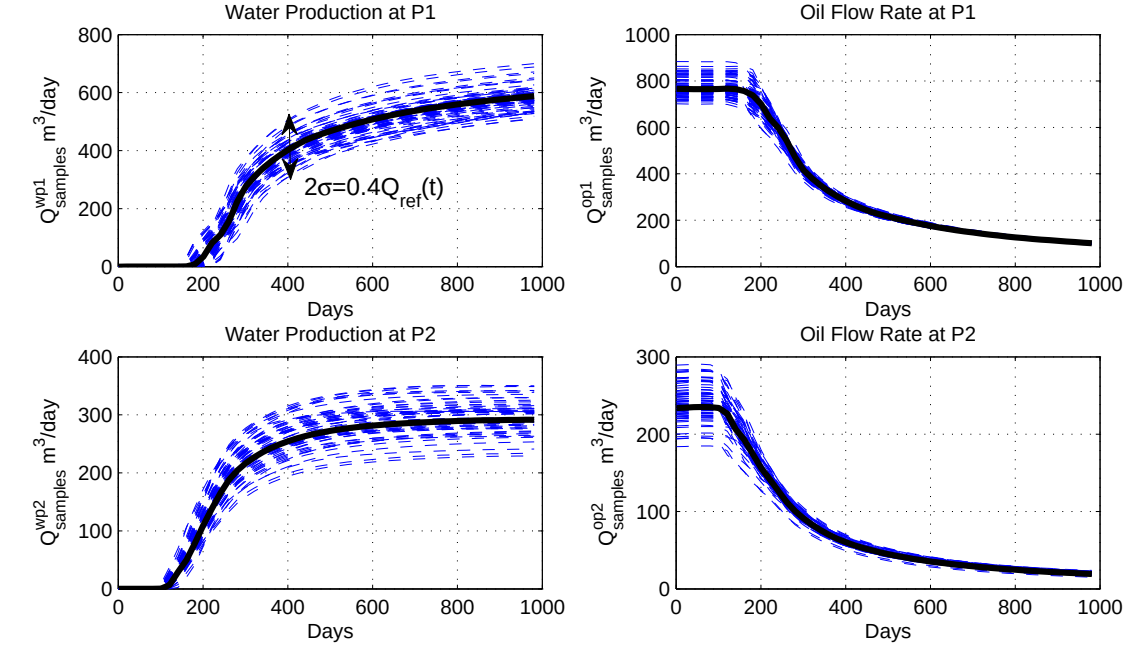

Figure 4.20: Full reservoir simulator runs for the samples from the posterior probability distribution of permeability when the observational error increases to $\sigma_D^2 = 0.2$, vs. the reference case history (the black solid curve). It demonstrates that when the observational error increases, the posterior samples have a larger mismatch with the reference case.

To assess the efficiency of method, let F denotes the running time for a full reservoir simulation and α denotes the running time for the polynomial chaos proxy where $F \gg \alpha$, then the computational complexity of the proposed MCMC method using the polynomial chaos proxy is $N \times F + n_1\alpha$, where N is the number of trial runs and n_1 is the number of MCMC steps to reach at the equilibrium state ($n \gg N$). The computation complexity of the traditional MCMC method is n_2F. Hence the speed-up ratio of the proposed method can be calculated as:

$$\text{Speed-Up} = \frac{n_2 \times F}{N \times F + n_1\alpha} \approx \frac{n_2}{N} \tag{4.67}$$

When the number of dimension increases, the number of the MCMC steps required to reach at the equilibrium state grows exponentially and n_2 becomes much larger than the number of trial runs to construct a reliable polynomial chaos proxy. Hence the proposed method presents a significant speedup.

For this synthetic reservoir case study, the full reservoir simulator is implemented using MATLAB on a local workstation. The running time for a full reservoir simulation is $F = 185s$ while the pseudo-simulator using the polynomial chaos proxy along with the Karhunen-Loeve decomposition method has a running time of $\alpha = 3s$. When four eigenvectors are retained, the number of the traditional MCMC steps to reach the stationary distribution using the full reservoir simulator is $n_2 = 10810$ while the number of the required steps for MCMC using the polynomial chaos proxy to reach at the equilibrium state is $n_1 = 8025$. Hence, the speed-up is $\frac{10810\times185}{(185000+8025\times3)} = 9.5$ times, when four eigenvectors are retained. When the number of retained eigenvectors increases, the number of the required full reservoir simulation runs for the convergence of the MCMC method increases drastically and it becomes practically infeasible. For real oil fields, where million cells may be involved, typically more than 10^7 reservoir simulation might be required for the traditional MCMC method to converge to the posterior distribution. Accordingly, the speed-up that the proposed method renders will be raised to million times!

In the next case study, we consider a remarkably nonlinear case whose misfit surface has two global minimums. As a result the high-order terms in the misfit formulation of the equation (4.64) becomes prominent in the posterior distribution of the permeability field.

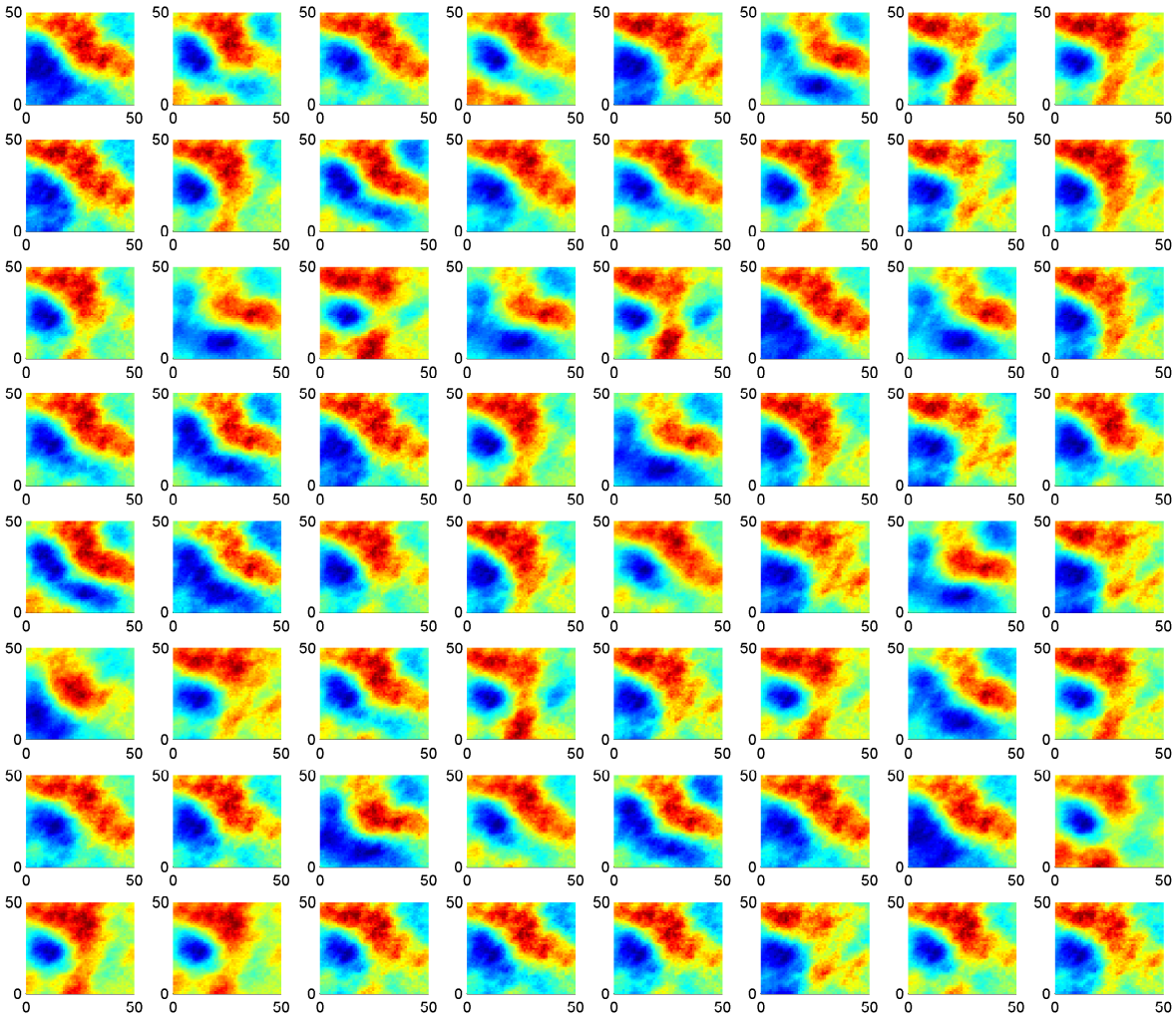

Figure 4.21: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.1$ and 4 eigenvectors are retained. PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 as the reservoir proxy model.

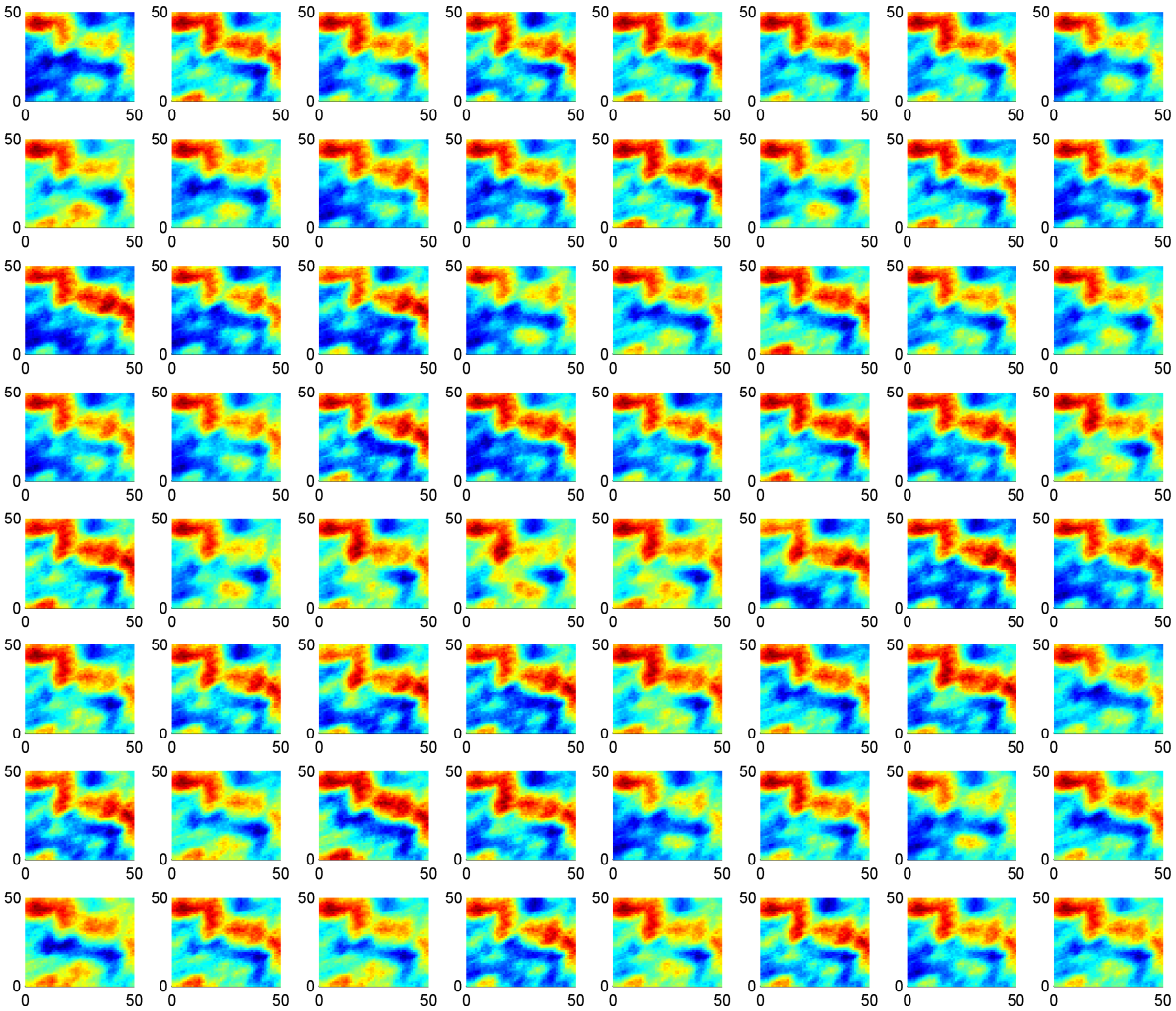

Figure 4.22: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.1$ and 40 eigenvectors are retained. PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 as the reservoir proxy model.

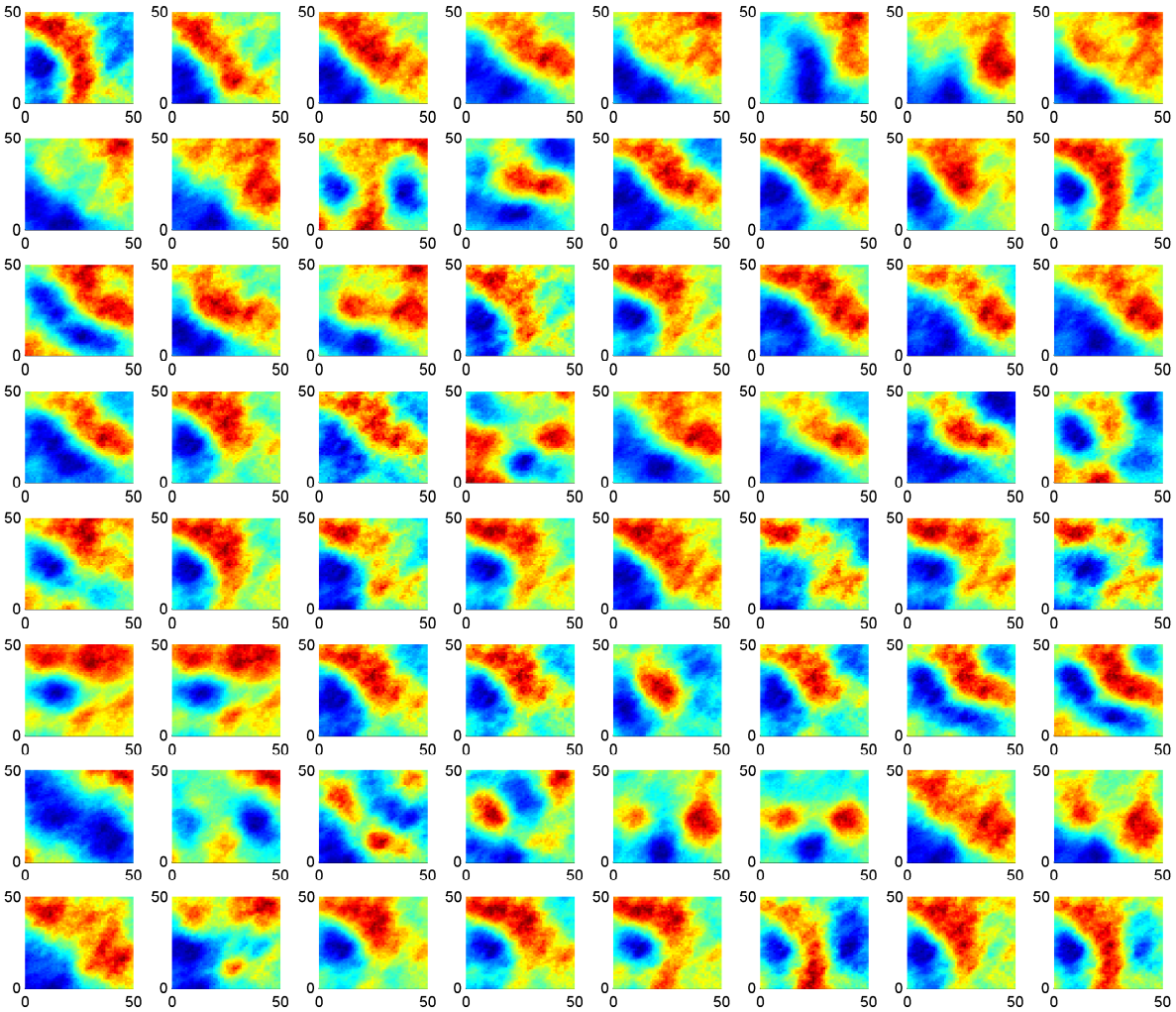

Figure 4.23: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.2$ and 4 eigenvectors are retained. PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 as the reservoir proxy model.

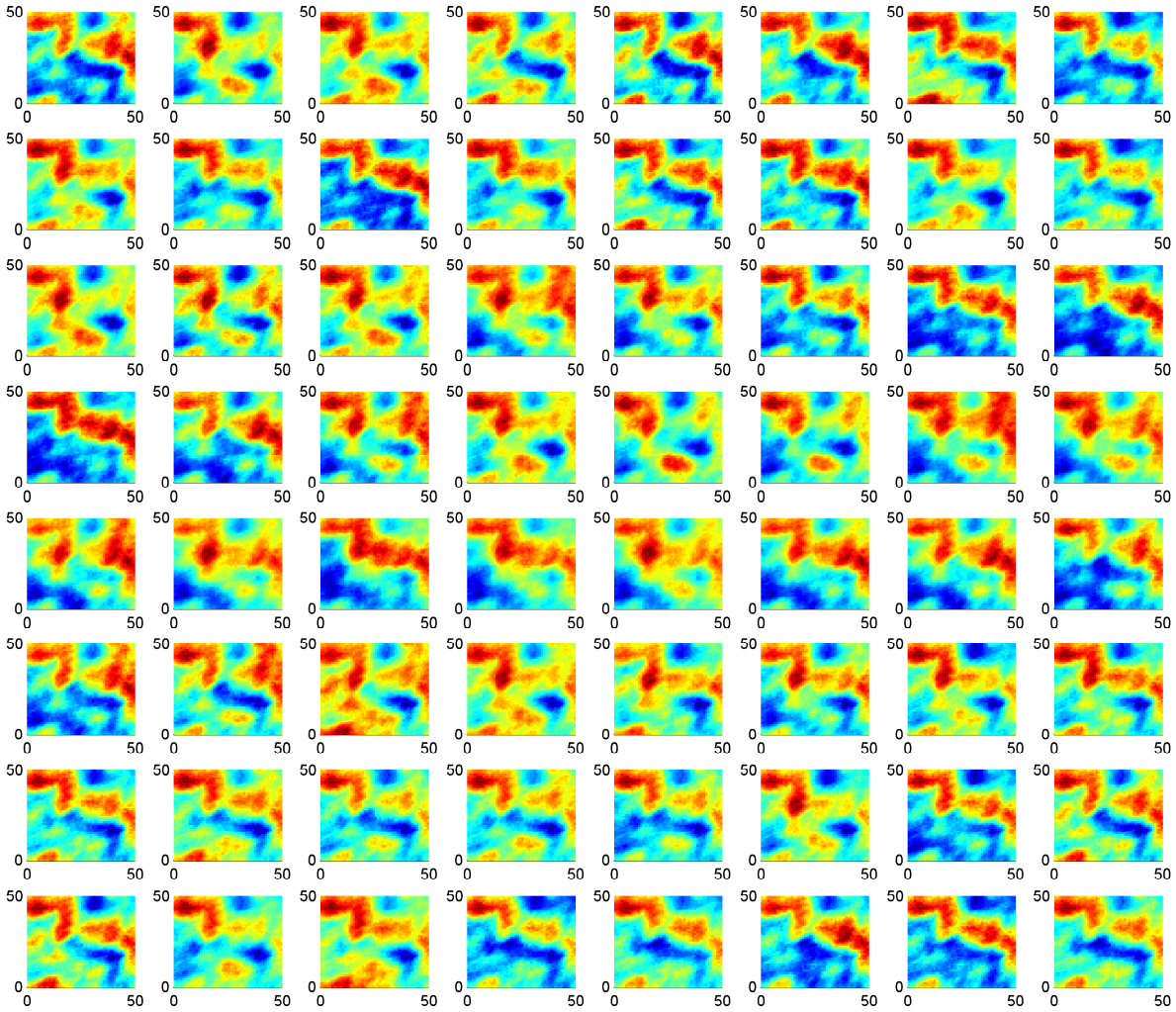

Figure 4.24: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.2$ and 40 eigenvectors are retained. PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 as the reservoir proxy model.

Case II

The second synthetic case model represents a 2D horizontal square reservoir with two water injectors in the left corner and only one producer in the right as illustrated in Figure 4.25. The reservoir covers an area of $500 \times 500 m^2$ and has a thickness of $10\,m$. It is discretized into a 50×50 horizontal 2D grid. Similar to the previous case study, it is essentially an incompressible two-phase unit mobility oil-water system with zero connate water saturation and zero oil residual system. The porosity is assumed to be 0.3 all over the field. The model is run for 1000 days with the injectors under rate control and the producers under bottom whole pressure (BHP) control with predefined rates and BHP. The injection rate is $500\,\frac{m^3}{day}$ for each injectors. The observation is the oil production rate for 1000 days. Compared to case study I where we had more information (oil and water rates for two producers), the problem is more uncertain here.

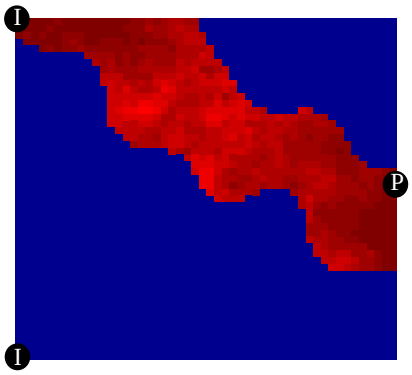

Figure 4.25: The synthetic water flooding model with two injectors on the left and one producer on the right. It also shows the reference permeability field used to generate the observed data of the oil flow rate.

The prior geological knowledge indicates that the permeability field is a fluvial channelized reservoir, which has a fine sand texture with permeability $100\,mD$ and the permeability of the background matrix is $10\,mD$. The contrast between the channel and the matrix is $\frac{\overline{K}_{high}}{\overline{K}_{low}} = 10$. We also assume the variance of the permeability field is as small as $1\,mD$. The relative sinuosity of the fluvial channel varies from 0.1 to 0.3, the thickness from $50\,m$ to $200\,m$, the amplitude from $100\,m$ to $200\,m$, and the wavelength ranges between $200\,m$ and $250\,m$. No other information about the orientation and the location of the channel is available. The objective is to obtain samples from the posterior distribution of the permeability field based on the observed data for the oil flow rate for 1000 days.

Figure 4.26 illustrates the saturation map of the water flooding process of the fluvial channel for 1000 days.

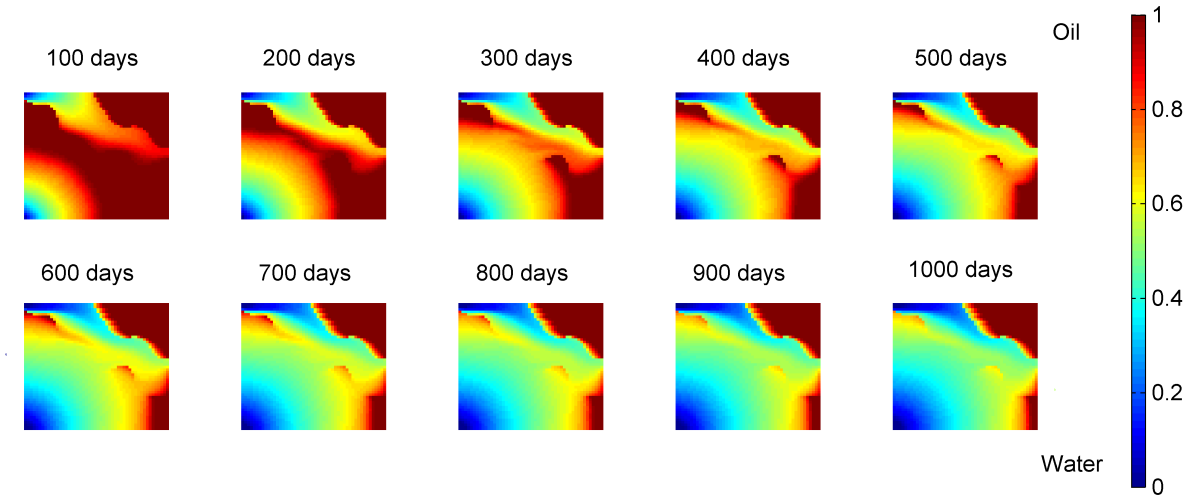

Figure 4.26: The water flooding process of the fluvial channel case II for 1000 days.

Figure 4.27 shows the observed data of oil flow rate before adding noise. The true permeability model which is used as the reference case to generate the production data is a fluvial channel extended from one injectors to the producer as shown in Figure 4.25. Obviously the identical channel from the other injector to the producer gives the same oil flow rates as the reference case. Hence, at least two permeability models give the misfit surface zero value. Consequently, the posterior distribution of ξ in the equation (4.65) will essentially be considerably nonlinear. To assess the uncertainty associated with the permeability field, using the polynomial chaos proxy, we follow the same procedure of the previous case study while the higher order terms of the polynomial chaos proxy will be prominent in the Bayesian formulation (4.65).

- **Reduced-order parameterization of the permeability field**
  Using the prior information about the fluvial channel, we generate 1000 realizations and compute the ensemble mean and the covariance matrix. Since the orientation and the location of the fluvial channel are not specified in the prior geological knowledge, the realizations have a wide range of variance. We make the problem more challenging by assuming that the prior parameterization of the permeability field by the fluvial channel property is not available. Hence, the starting point is to parameterize the permeability field based on the given $N = 1000$ realizations. We perform the Karhunen-Loeve parameterization (linear PCA) using

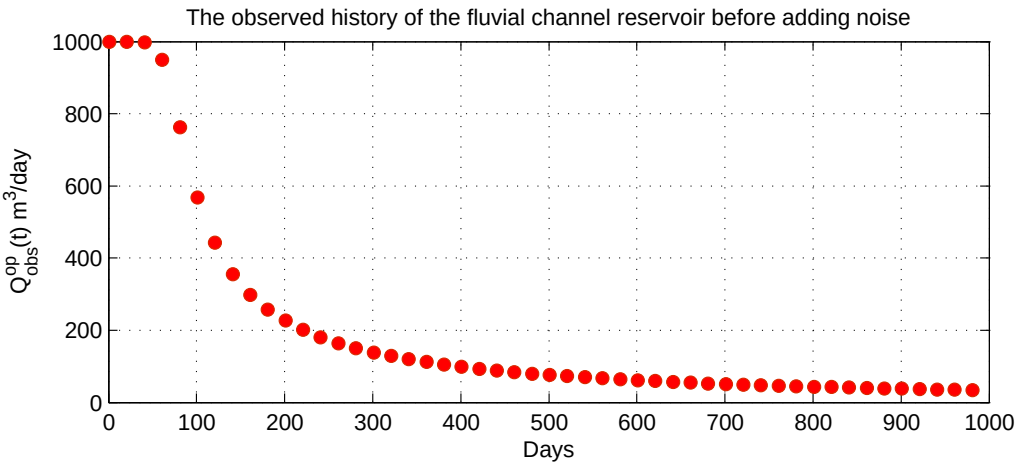

Figure 4.27: The observed history of the reservoir (before adding noise)

the ensemble covariance matrix and discard the small eigenvalues to reduce the dimension.

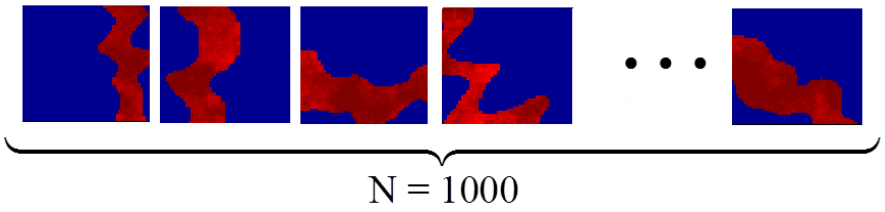

Figure 4.28: To train the polynomial chaos proxy, **N** realizations are generated based on the given prior distribution.

Figure 4.29 shows the scree graph of the eigenvalues. The decay rate of eigenvalues is smaller than the previous case study as the variance of the realizations is higher. For a good sampling, the minimum of 40 eigenvector has to be retained to preserves the shape of the channel. Discarding the small eigenvalues will smooth the sharp transition (high frequency changes) between the two distinct textures.

- **Construct the polynomial chaos proxy**

  To train the polynomial chaos proxy with 40 random variables and $N = 1000$ simulation runs, we obtain the reduced-terms polynomial chaos expansion of order 8 for oil flow rate:

$$\mathbf{Q^{op}}(\xi, \mathbf{t}) = Q_0^{op} + \sum_{i=1}^{P} Q_{i,1}^{op}\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op}(\xi_i^2 - 1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} Q_{ij,2}^{op}(\xi_i\xi_j) + \dots \quad (4.68)$$

  As discussed earlier, the reduced-terms polynomial chaos proxy preservers the relevant terms based on the impact factor. The cut-off value of $\epsilon = 0.1$ is chosen

to drop the insignificant terms. To find the coefficients of the reduced-terms polynomial chaos expansion, we use regression-based PCM that minimizes the sum of squared errors.

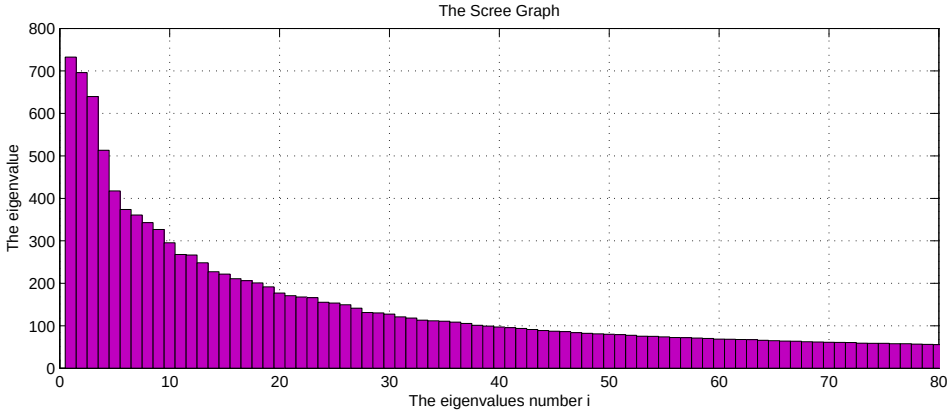

Figure 4.29: The Scree plot for the ensemble covariance matrix of 1000 realizations

- **Bayesian inference formulation**

  Using misfit formulation of (4.26) and plugging in the polynomial chaos proxy substitute of the equation (4.68) for oil flow rate, we obtain:

$$\begin{aligned}\mathbf{S}(\xi_{\mathbf{1}}, \xi_{\mathbf{2}}, \ldots, \xi_{\mathbf{p}}) &= \sum_{i=1}^{M} \frac{(\mathbf{Q}^{\mathbf{op}}_{\mathbf{proxy}}(\mathbf{t_i}) - \mathbf{Q}^{\mathbf{op}}_{\mathbf{obs}}(\mathbf{t_i}))^2}{\sigma^2(t_i)} \\ &= S_0 + \sum_{i=1}^{p} S_{i,1}\xi_i + \sum_{i=1}^{p} S_{ii,2}(\xi_i^2 - 1) + \sum_{i=1}^{p-1}\sum_{j>i}^{p} S_{ij,2}(\xi_i\xi_j) + \ldots.\end{aligned} \tag{4.69}$$

  We define σ similar to the previous case study, $\sigma(t) = \sigma_D\, Q^{op}_{obs}(t)$, where σ_D is constant. We solve the problem for $\sigma_D^2 = 0.1$ and $\sigma_D^2 = 0.2$. We also assume that the minimum noise (σ_{min}) is 30 $\frac{m^3}{day}$. The equation (4.65) gives an analytical approximation for the posterior probability distribution. We also assume the priors for ξ_i are all standard Gaussian. The joint distribution $\rho_\xi(\xi)$ is already presented in the equation (4.66).

- **Metropolis-Hasting MCMC**

  Having an analytical expression for the posterior distribution, we apply the Metropolis-Hasting MCMC, with a multivariate Gaussian as the proposal distribution, to sample from the posterior distribution of the permeability field. To check the

convergence of the chain, all moments up to the 8^{th} have to be examined for our desired precision (the 8^{th} order of the polynomial chaos expansion is used).

To decide on the order of the polynomial chaos proxy, given N realizations and p random variables ($N > p$), the following algorithm is implemented:

1. All the first order terms are retained, $\{\xi_1, \xi_2, \ldots, \xi_p\}$
2. The order of the polynomial chaos is increased to the next order, $PC_m(\xi, t)$
3. The insignificant terms in $PC_m(\xi, t)$ are dropped based on their corresponding impact factor and the pre-determined cut-off value ϵ
4. While the total number of the unknown coefficients in the reduced-terms polynomial chaos proxy is smaller than N, repeat the procedure.

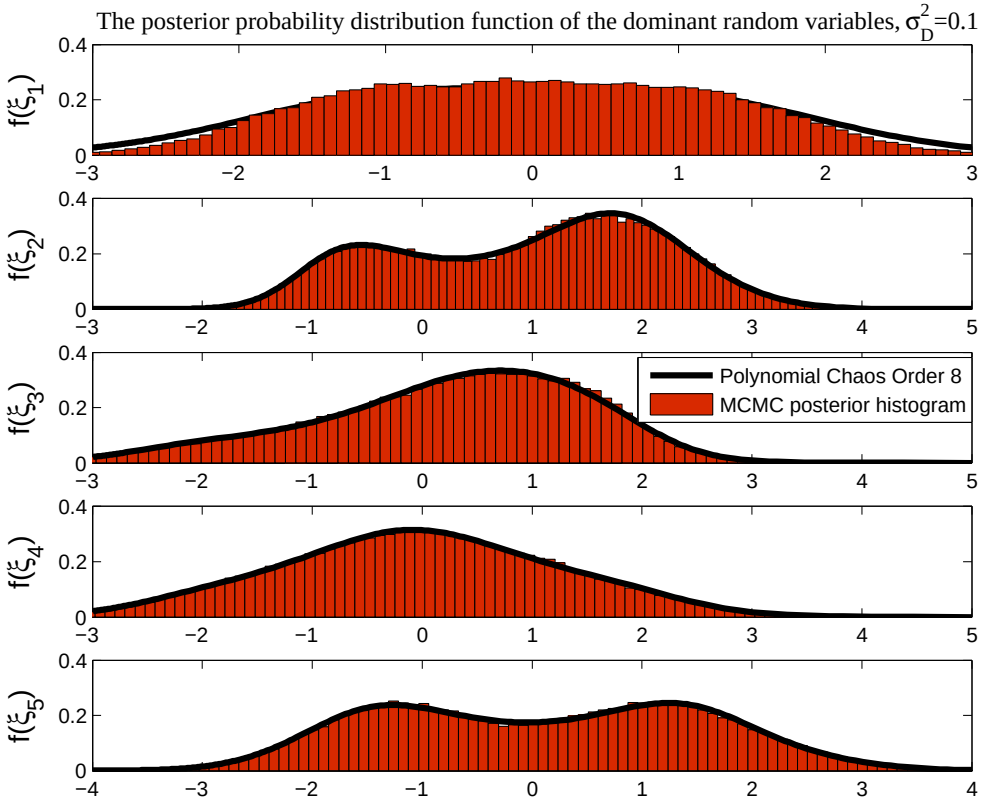

Figure 4.30: Estimation of the marginal posterior distribution of the most dominant random variables $\{\xi_1, \xi_2, \ldots, \xi_5\}$ using the 8^{th} order polynomial chaos proxy, compared to the MCMC result with reservoir simulator.

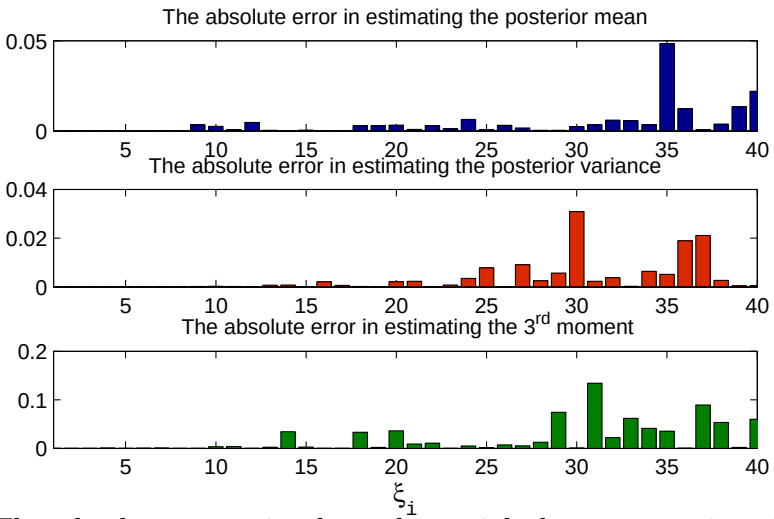

Figure 4.31: The absolute error in the polynomial chaos approximation of the first, second and third moments compared to the Monte Carlo results when $\sigma_D^2 = 0.1$.

Figure 4.30 shows the marginal posterior distribution for the 5 most dominant random variables $\{\xi_1, \xi_2, \ldots, \xi_5\}$. The result of the polynomial chaos proxy of order 8 is compared with the expensive MCMC result using full-reservoir simulation. The dimensionless error in the observation is $\sigma_D^2 = 0.1$. The histogram shows the true distribution (MCMC using reservoir simulator) where the number of bins is 100. Figure 4.30 proves a perfect match between the polynomial chaos proxy of order 8 and the true posterior distribution.

For the random variables associated with the smaller eigenvalues (higher order random variables), more than 10^7 full reservoir simulation is required for the MCMC method to converge, which is not practically feasible. Hence we compare the accuracy of the polynomial chaos proxy in the approximation of the first, second and third moments of the random variables with the Monte Carlo method. Figure 4.31 illustrates the absolute error in the approximation of the first three moments. It demonstrates that the accuracy decrease for the higher frequency modes, because in the reduced terms polynomial chaos expansion, we have neglected some of the terms associated with them.

When five eigenvectors are retained, the number of the traditional MCMC steps to reach the stationary distribution using the full reservoir simulator is $n_2 = 15205$ while the number of the required steps for MCMC using the polynomial chaos proxy to reach at the equilibrium state is $n_1 = 10794$. Hence, the speed-up of the proposed method compared (using Karhunen-Loeve decomposition method) to the traditional MCMC method is $\frac{15205 \times 185}{(185000 + 10794 \times 3)} = 12.94$ times, when five eigenvectors are retained.

To study the convergence of the analytical formulation for the posterior distribution using the polynomial chaos proxy to the true posterior distribution, we construct different orders of the polynomial chaos proxy and compare it against the true distribution. Figure 4.32 shows the marginal posterior distribution of ξ_2 for different orders of the reduced-terms polynomial chaos proxy versus the true posterior distribution. It demonstrates the convergence of the polynomial chaos proxy-based formulation for the posterior distribution to the true distribution as the order of the expansion increases. The two maximum points in the posterior distribution of ξ_2 corresponds to the two different solution sets; the channels extended from the upper injector to the producer (as the reference case), and alternatively the channels elongated from the lower injector to the producer.

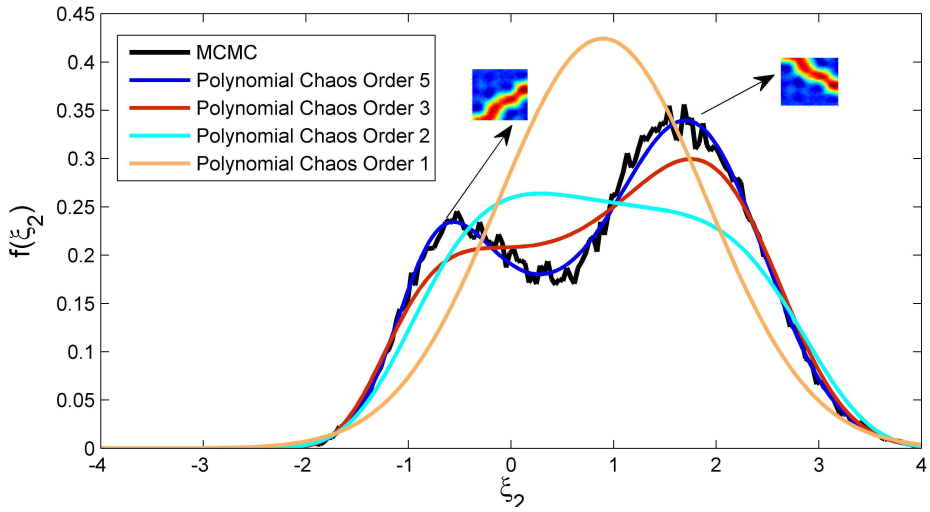

Figure 4.32: The polynomial chaos proxy-based estimation for the posterior distribution of ξ_2 converges to the true distribution as the order of the expansion increases.

Accordingly the high-order terms in the misfit formulation of (4.26) become prominent in the estimation of the posterior distribution. This verifies one of the great advantages of the polynomial chaos proxy over other proxy models that it guarantees the convergence in distribution as the order of the expansion increases.

The first order terms (linear terms) in the polynomial chaos proxy estimates the posterior probability density function with a Gaussian distribution. The higher-order terms captures the non-Gaussian shape of the residual systematically. Hence the polynomial chaos proxy can be thought as a generalization to efficient linear estimators (such as the Ensemble Kalman Filter [49, 1]) that works for the non-Gaussian posterior distribution as well.

Since we used the linear PCA to parameterize the permeability field, we can obtain samples from the posterior distribution of the permeability field. Figure 4.33 and Figure 4.34 show different samples from the posterior distribution of the permeability field when the dimensionless error in the observation is $\sigma_D^2 = 0.1$ and $\sigma_D^2 = 0.2$. Figure 4.35 and Figure 4.36 illustrate the full-simulator run for the obtained samples from the posterior distribution of permeability field when $\sigma_D^2 = 0.1$ and $\sigma_D^2 = 0.2$.

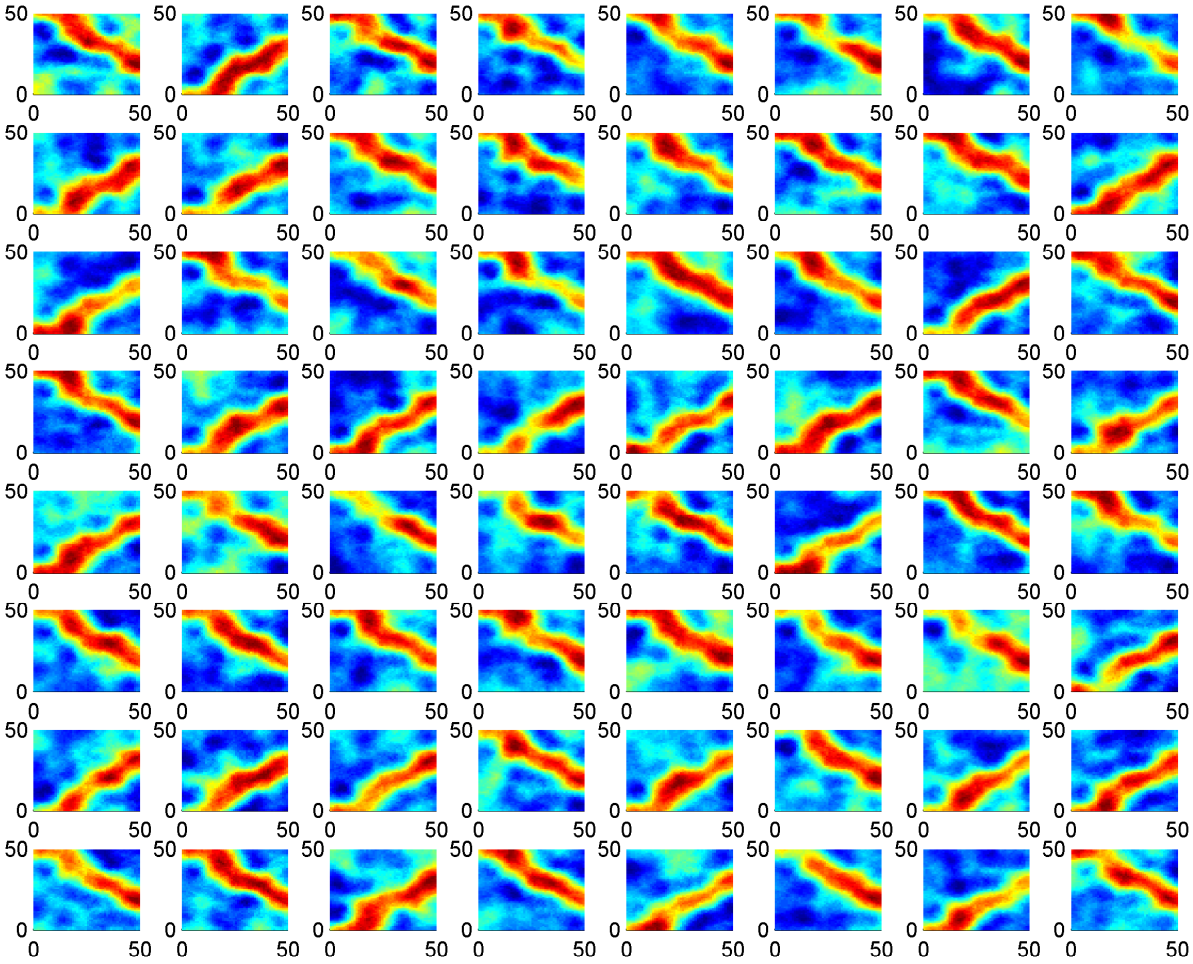

Figure 4.33: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.1$ and 40 eigenvectors are retained. Linear PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 is used for the reservoir proxy model.

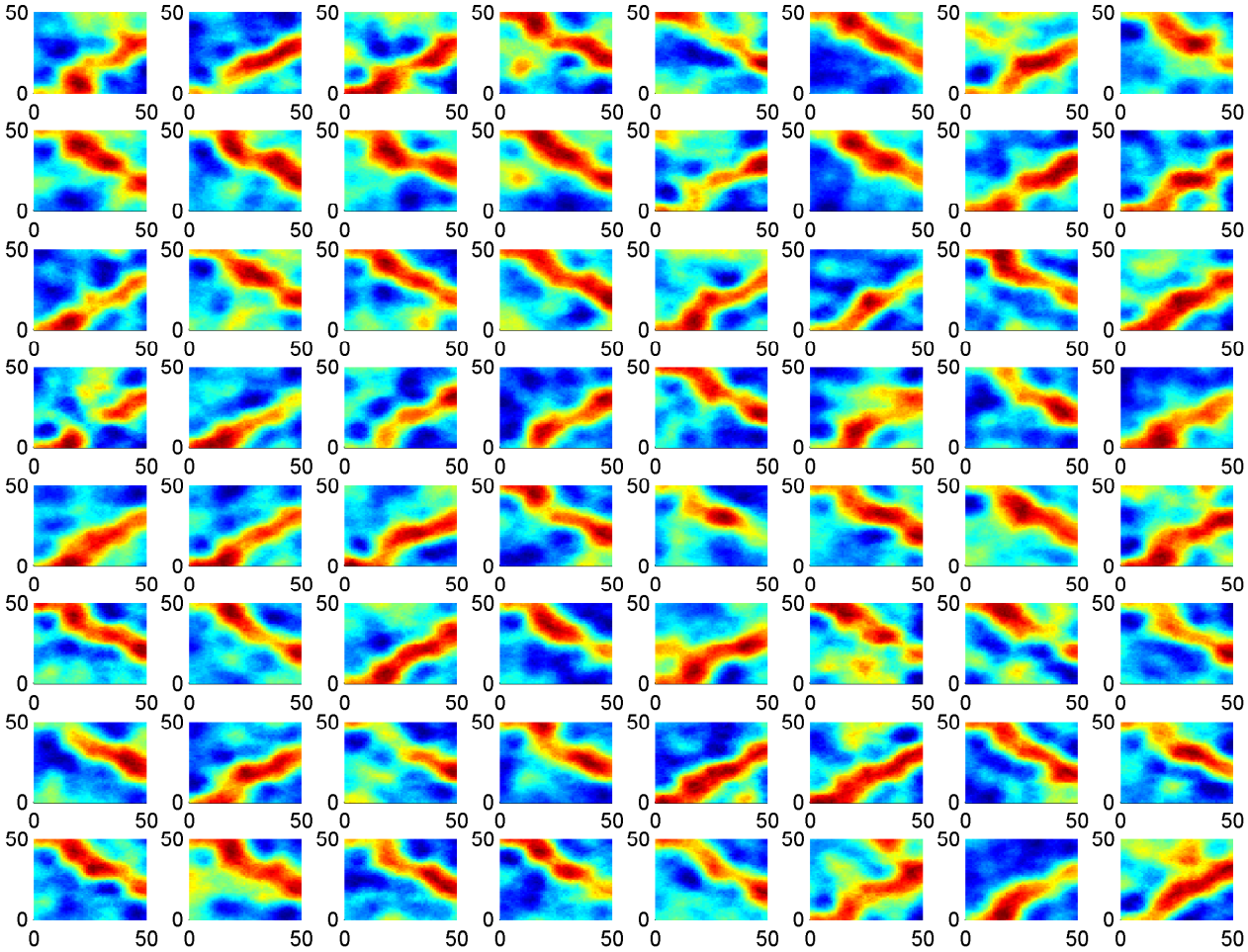

Figure 4.34: Samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.2$ and 40 eigenvectors are retained. Linear PCA is used for dimensionality reduction and the reduced-terms polynomial chaos expansion of order 8 is used for the reservoir proxy model.

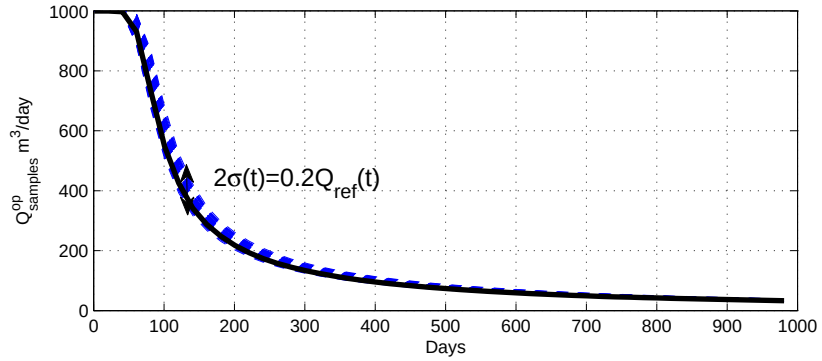

Figure 4.35: Full reservoir simulator runs for the samples from the posterior probability distribution of permeability when the dimensionless error in the observation is $\sigma_D^2 = 0.1$, vs. the reference case history (the black solid curve).

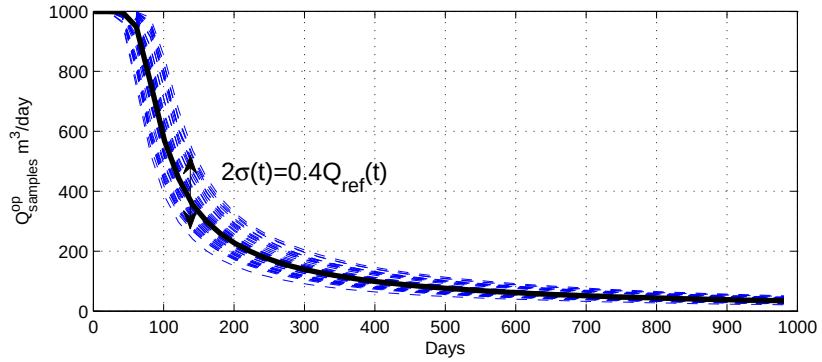

Figure 4.36: Full reservoir simulator runs for the samples from the posterior probability distribution of permeability when the observational error increases to $\sigma_D^2 = 0.2$, vs. the reference case history (the black solid curve). It demonstrates that when the observational error increases, the posterior samples have a larger mismatch with the reference case.

This case study verifies the efficiency of the polynomial chaos proxy framework to sample from the posterior distribution when the forward function is remarkably nonlinear. We proved that the high-order polynomial chaos proxy can capture the nonlinear effects of the forward model on the posterior distribution.

4.5 Summary

A novel polynomial chaos proxy-based history matching and uncertainty quantification method is presented in this chapter that can be employed to solve complex geological structures in inverse problems. Modern reservoir management has a paramount focus on predicting the likely range of field recoveries and consequently providing economic evaluations of different field development strategies. To estimate possible ranges in recovery factors, we need enough samples from the posterior probability distribution of geological structure. For complex geological structures, more than 10^6 full reservoir simulation runs might be required to accurately sample from the posterior probability distribution. This is impracticable for high-resolution geological models when there are many unknown geological parameters. The proposed algorithm can efficiently sample from the posterior probability distribution of a complex geological structure by running a surrogate model constructed from a few hundred runs of the full reservoir simulation. The proposed surrogate-accelerated algorithm is composed of four main components:

- Dimensionality reduction
- Reduced-terms polynomial chaos expansion
- Bayesian Inference
- Metropolis-Hasting MCMC

Dimensionality reduction techniques and the polynomial chaos proxy were discussed in detail in the previous chapters. To mitigate the curse of dimensionality, we proposed the reduced-terms polynomial chaos expansion in this chapter which preserves only the relevant terms in the polynomial chaos representation. We demonstrated that the sparsity pattern in the polynomial chaos expansion, when used with the Karhunen-Loeve decomposition method or kernel PCA, can be systematically captured. We introduced the impact factor for each term in the polynomial chaos expansion to preserve only the relevant terms.

Bayesian inference basically represents learning from the observed data. It formulates the posterior probability distribution based on a given prior assumption and forward model. The forward model in the likelihood function is approximated here with the polynomial chaos expansion, instead of a full reservoir simulation. Convergence of the forward approximation (the polynomial chaos proxy) implies convergence of the posterior distribution. To estimate the possible range in recovery factors and reservoir

management, we draw samples from the posterior distribution computed by the polynomial chaos proxy. Metropolis-Hasting MCMC was employed to accurately sample from the posterior distribution. Since running the polynomial chaos proxy is very cheap, the millions of samples required for the Metropolis-Hasting MCMC to reach its equilibrium state can be efficiently acquired.

For future development, other MCMC methods can be examined to explore the posterior distribution. Since the polynomial chaos proxy is prone to have a considerable amount of error for highly improbable realizations, an adaptive and iterative construction of the polynomial chaos proxy can also be investigated. The proposed surrogate-accelerated Bayesian inverse analysis can be coherently used in practical reservoir optimization workflows and uncertainty assessments.

Chapter 5

Cluster Based Bayesian Inference using Polynomial Chaos Proxy

5.1 Introduction

In Chapter 2 the polynomial chaos expansion was discussed as the span of Hermite polynomial functionals of a Gaussian process. According to the Cameron-Martin theorem [25], the Fourier Hermite series converge to any L_2 functional in the L_2 sense. In the context of stochastic processes, this implies that the polynomial chaos expansion converges to any processes with finite second-order moments. Therefore, a stochastic process can be represented in terms of Hermite orthogonal polynomials. Although for any arbitrary random process with finite second-order moments, the Hermite-Chaos expansion converges in probability, the Cameron-Martin theorem proves that the convergence rate is optimal for Gaussian processes and the error decays exponentially as the order of the polynomial chaos increases [25]. However, for other types of stochastic processes the convergence rate may be substantially slower [199]. To resolve this issue for a certain type of stochastic processes, Askey-scheme of polynomials could be used to construct the orthogonal polynomials of chaos expansion [199]. The Askey-scheme consists of a set of hypergeometric orthogonal polynomials that satisfies certain types of differential equations [199]. Hermite polynomials are a subset of the Askey-scheme. Each subset of the orthogonal polynomials in the Askey-scheme has different weighting. Xiu and Karniadakis [199] demonstrated numerically that the optimal (exponential) convergence rate can be achieved with each subset of the Askey polynomial chaos expansion for their corresponding stochastic processes by solving a stochastic ordinary

differential equation, for which the exact solutions can be obtained. To increase the convergence rate of polynomial chaos with stochastic inputs of arbitrary probability measures, the *multi-element generalized polynomial chaos* method was proposed which adaptively constructs the polynomial chaos basis [187].

In high-dimensional problems, the multi-element generalized polynomial chaos for an arbitrary probability measure is hard to implement [187]. However, for a specific class of problems that the random inputs can be effectively clustered into different scenarios where each cluster share the similar statistics, the optimal convergence rate can be efficiently achieved. It is due to the fact that effective clustering gives a better representation of a random field. For instance, the distribution of heights among people is better modeled by considering men and women as separate clusters, and consequently different parameters can be estimated more accurately if one separates out subpopulations. Moreover, when each cluster can be reasonably approximated by a Gaussian model, the standard Wiener polynomial chaos expansion will efficiently estimate the random output distribution as the exponential convergence rate is guaranteed by the Cameron-Martin Theorem.

Classification is a common practice in various fields of science. As Kendall said, *"one of the basic problems of science in reducing the world to order (or, if you prefer it, in imposing a manmade order on the complexity of things) is to classify"* [92, 46]. In general, classification can be defined as the clustering of objects based on their similar characteristics. In statistics, cluster analysis is considered in two different situations; supervised learning and unsupervised learning. In supervised learning, there is prior information that can be used to obtain further information on the group structure [46]. For example different realizations of a reservoir can be clustered into different groups by geological interpretation. In unsupervised learning, there is no prior information on the grouping of the data and statistical tools (e.g. multidimensional scaling [40, 95]) are used for classification [46].

In his dictionary of statistics, Everitt defines cluster analysis as, *'a set of methods for constructing a hopefully sensible and informative classification of an initially unclassified set of data, using the variable values observed on each individual. All such methods essentially try to imitate what the eye-brain system does so well in two dimensions'* [50]. For instance, in Figure 5.1, four distinct groups can be identified by eye. In high-dimensional problems however, and even in some complex two dimensional cases, discovering an order in the structure of data is more involved and requires advanced clustering techniques [46].

For cluster analysis, various methods have been practiced in literature [46, 91, 3].

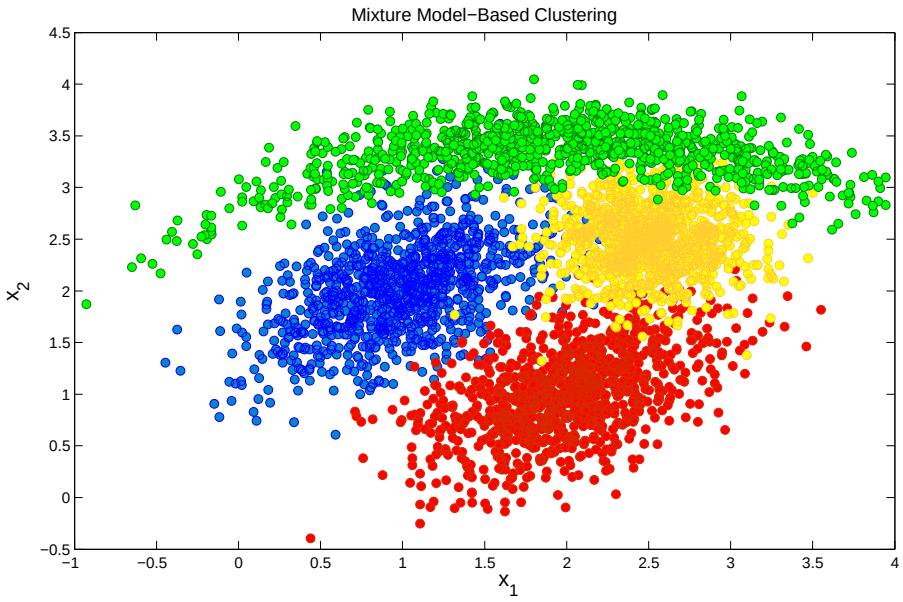

Figure 5.1: Clustering of 2-dimensional data into four different populations

These methods generally vary from mainly heuristic methods to methods based on statistical techniques. For example, one popular algorithm is hierarchical clustering, where based on an optimality criterion at each step two clusters are either merged (agglomerative) or divided (divisive) [83, 46]. Another well-known technique is K-mean algorithm that relocates the observation between a predetermined number of clusters [178, 46]. The K-mean algorithm clusters data into K different groups by minimizing the total mean square error between the training samples and their representative cluster centroid [178]. Both of these methods require either a prior knowledge on the number of the clusters or a pre-determined optimality criterion to optimize the number of the clusters [46]. Essentially, all the cluster analysis techniques have been developed to determine the number and structures of the clusters [46]. These entail the determination of

1. similarity and separation of clusters,
2. shape (distribution) of underlying clusters,
3. relative sizes and compactness of clusters [46].

Mixture modeling is a probabilistic approach to clustering where dataset are assumed to be rooted from a mixture of clusters with different probability density functions [175]. Then each individual sample does not exclusively belongs to a specific clus-

ter. In other words, each cluster can contain a sample with probability $P_{\mathcal{M}_i}$. Hence, in the mixture modeling approach the issue of cluster analysis reduces to a single concern, that of model selection [175]. Considering that each combination of clusters results in a different overall distribution for the data, cluster analysis essentially concerns the problem of comparison among the number of possible models and the probability density function of the underlying clusters [175].

There are usually trade-offs between the number of clusters and the complexity of the underlying models. If a simpler model is used, then more clusters may be need to provide an accurate representation of the data. If a more complex model is used, then data can be represented with a few clusters [46]. For example, to fit data from a single Gaussian cluster whose covariance is a single elongated ellipsoid, one requires more than one hyper-spherical Gaussian cluster.

The objectives of clustering in our work are:

1. to make the problem less nonlinear under each cluster. Hence, the polynomial chaos proxy converges faster for each cluster and the overall posterior distribution can be estimated more efficiently using different proxies.

2. to compare the possibility of different scenarios given data.

Our approach to cluster analysis in mixture modeling is based on the posterior probabilities of models, known as the "Bayesian Model Selection". In statistics, the Bayesian model selection was first introduced in the work of Jeffreys in 1939 [82] and pioneered by Cox [35], Akaike [2], Schwarz [162]. A rich review on the Bayesian model selection is presented by Clyde and George [31]. To illustrate the Bayesian model selection consider models $M_1, M_2, \ldots, M_K$ with prior probabilities $p(M_j),\ j = 1, \ldots, K$ (often assumed to be equal), then by the Bayes's rule, the posterior probability of model M_j given data D is proportional to the probability of the data given model M_j, times the model's prior probability, namely

$$p(M_j|D) \propto p(D|M_j)p(M_j). \tag{5.1}$$

For the posterior calculation, the Markov Chains Monte Carlo has been successfully practiced in the Bayesian framework [6, 5, 30, 26, 68]. Ratery et al. examined application of the Bayesian model selection in five areas; genetics, sports, ecology, sociology, psychology [89, 140, 60]. Subsequently, in the reservoir modeling context, Gallagher et al. applied MCMC for the Bayesian inference to select the optimal reservoir model [58]. In large problems, the MCMC method may need more than 10^6 expensive runs of reservoir simulation and is not practically feasible. Our solution is to employ the polynomial

chaos proxy to efficiently estimate the posterior probability density function under each cluster and calculate the posterior probability of each model $p(M_j|D)$.

This chapter is structured as follows; first we review the mixture modeling approach for clustering, followed by an introductory note on the Bayesian model selection. Then we propose the cluster based polynomial chaos framework for the Bayesian inference. Through examples we will show that the cluster-based polynomial chaos proxy is more efficient in approximating the mixture posterior distribution comparing to the standard non-cluster based polynomial chaos proxy. Finally, we apply the cluster-based polynomial chaos proxy to calculate the posterior probability of three different scenarios for the second layer of the Stanford VI reservoir.

5.2 Mixture Modeling

Mixture modeling is a probabilistic approach to represent the presence of clusters with different probability distribution within an overall distribution of data [112, 51]. Hence, the clusters for a $p-$dimensional dataset with N observations are considered to have different populations while data in each cluster has the same statistics. Accordingly, the data are assumed to root from a mixture of K underlying populations, each corresponding to a cluster with a specific probability density function. Hence, the clustering analysis is to determine

- the probability density function in each cluster,
- the number of clusters,
- an optimization algorithm, and
- criteria to select the optimal method [56, 105].

Since there are various optimization methods to the parameter estimation problem as well as numerous models to estimate the probability density function of each cluster, there is a significant amount of opportunity available in the development of the clustering method [56, 105, 112, 51].

In this work, we assume the number of clusters is already determined either by the geological information or the statistical tools for the pattern recognition. Also, the prior distribution of the K underlying populations is considered to be multivariate Gaussian distribution with different statistics. The criteria to select the optimal method is the Bayesian model selection algorithm and the probability density function in each cluster is updated with the posterior probabilities given the observed data $\mathcal{D}$.

5.2.1 Gaussian Mixture Model-based Clustering

Karl Pearson [136], along with his other contributions to statistics, was the first mathematician to model a dataset resulting from two different populations as a mixture of Gaussian distributions. On the pursuit of his collaborative research with Raphael Weldon [193] on statistical analysis of Darwin's theory of evolution through *survival of the fittest*. They conducted an experiment of keeping thousands of crabs in two different environments for a period of time and measured the size of their carapaces [94]. Pearson proposed to fit a mixture of two Gaussian distributions to the crabs measurements and solve the parameter estimation problem to determine the death rates in relation to the laws of growth. However, their research did not advance further due to the complexity of the parameter estimation problem until the development of modern computational techniques.

In the Gaussian mixture model-based clustering, each population is assumed to have a multivariate Gaussian distribution $\mathcal{N}(\mu_k, \Sigma_k)$, where μ_k is the mean and Σ_k is the covariance matrix of the dataset in the k-th cluster. The mixture density is then given by

$$\begin{aligned} f(\mathbf{z}) &= \sum_{i=1}^{K} P(\mathcal{M}_i) g(\mathbf{z}|\mu_i, \Sigma_i) \\ g(\mathbf{z}|\mu_i, \Sigma_i) &= \mathcal{N}(\mu_i, \Sigma_i), \end{aligned} \tag{5.2}$$

where $P(\mathcal{M}_i)$ represents the probability of i-th component and consequently:

$$\sum_{i=1}^{K} P(\mathcal{M}_i) = 1. \tag{5.3}$$

We assume the prior parameters of each cluster (μ_i, Σ_i) are already known. The posterior probability density function of each cluster can be obtained using the Bayesian inference. In the next section, we explain the Bayesian model selection to compare the plausibility of different clusters.

5.3 Bayesian Model Selection

The Bayesian model selection is an alternative approach to the classical *hypothesis testing* [67]. Contrary to the classical likelihood-ratio test, the selection criteria in the Bayesian inference does not solely depend on any single set of parameters. It integrates over all possible parameters in each scenario with respect to their corresponding priors and gives the measure of how well a model is supported by observations.

Assume the realizations from the uncertain domain $\mathbf{z}$ is clustered into K different populations $(\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_K)$, where under each model class $P(\mathbf{z}|\mathcal{M}_i)$ represents the probability distribution of the uncertain parameters. The probability distribution of $P(\mathbf{z})$ can be written as the mixture probability density function of the underlying clusters;

$$f(\mathbf{z}) = \sum_{i=1}^{K} P(\mathcal{M}_i) f(\mathbf{z}|\mathcal{M}_i). \tag{5.4}$$

Let $P(\mathcal{D}|\mathcal{M}_i)$ be the probability distribution (likelihood) of the observed data under model class $\mathcal{M}_i$ and $P(\mathcal{M}_i)$ be the prior probability of model $\mathcal{M}_i$. The fundamental strategy in Bayesian model selection rest on analyzing the posterior probability distribution of models $P(\mathcal{M}_i|\mathcal{D})$. The posterior probability $P(\mathcal{M}_i|\mathcal{D})$ of model $\mathcal{M}_i$ given observed data $\mathcal{D}$ is given by the Bayes's theorem:

$$P(\mathcal{M}_i|\mathcal{D}) = \frac{P(\mathcal{M}_i)P(\mathcal{M}_i|\mathcal{D})}{\sum_K P(\mathcal{M}_i)P(\mathcal{D}|\mathcal{M}_i)}, \tag{5.5}$$

where

$$P(\mathcal{D}|\mathcal{M}_i) = \int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) f(\mathbf{z}|\mathcal{M}_i) d\mathbf{z} \tag{5.6}$$

is the marginal likelihood of $\mathcal{M}_i$. On the basis of observed data $\mathcal{D}$, the plausibility of two different model classes $\mathcal{M}_1$ and $\mathcal{M}_2$ is given by the Bayes factor:

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_1)}{P(\mathcal{D}|\mathcal{M}_2)} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \tag{5.7}$$

The Bayes factor updates the prior odds $\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$ to achieve the posterior odds $\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})}$. By replacing $P(\mathcal{D}|\mathcal{M}_i)$ with the marginal likelihood of equation (5.6), we obtain

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_1) f(\mathbf{z}|\mathcal{M}_1) d\mathbf{z}}{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_2) f(\mathbf{z}|\mathcal{M}_2) d\mathbf{z}} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}. \tag{5.8}$$

While Bayesian model selection is straightforward in principle, the calculation of the Bayes factor is not. Two challenges in the practical implementation of the Bayesian model selection is the calculation of the marginal likelihood of the equation (5.6) and the specification of priors. For the selection of priors one simple and popular choice is the uniform prior, where

$$P(\mathcal{M}_1) = P(\mathcal{M}_2) = \ldots = P(\mathcal{M}_K) = \frac{1}{K}. \tag{5.9}$$

The calculation of the marginal likelihood $(\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_1) f(\mathbf{z}|\mathcal{M}_1) d\mathbf{z})$ often involves multidimensional integrals that are hard to evaluate and pose a major challenge to the

Bayesian model selection. When the exact calculation of the marginal likelihood is not feasible, the Markov Chains Monte Carlo methods are commonly used to estimate multi-dimensional integrals and evaluate the Bayes factor [107, 26, 65, 58]. Several other schemes have also been developed by Bayesian authors for the numerical calculation of the Bayes factor. In particular, the reversible jump approach [68, 147] and path sampling [4] have gained popularity in recent years.

In high-dimensional problems, one popular approximation for the $P(\mathcal{D}|\mathcal{M}_j)$, when $h(\mathbf{z}) = \log\big(f(\mathcal{D}|\mathbf{z},\mathcal{M}_j)f(\mathbf{z}|\mathcal{M}_j)\big)$ is sufficiently smooth and well-behaved, is obtained by the Laplace's method as

$$P(\mathcal{D}|\mathcal{M}_j) \approx (2\pi)^{\frac{d}{2}}|H(\tilde{\mathbf{z}})|^{\frac{1}{2}}f(\mathcal{D}|\tilde{\mathbf{z}},\mathcal{M}_j)f(\tilde{\mathbf{z}}|\mathcal{M}_j) \tag{5.10}$$

where d is the dimension of uncertain domain $\mathbf{z}$, $\tilde{\mathbf{z}}$ is the maximum likelihood of $h(\mathbf{z})$ and $H(\tilde{\mathbf{z}})$ is minus the inverse Hessian of $h(\mathbf{z})$ evaluated at the maximum likelihood point [180, 30]. The approximation is obtained by replacing the Taylor series around the maximum likelihood point $h(\mathbf{z}) = h(\tilde{\mathbf{z}}) - \frac{1}{2}(\mathbf{z}-\tilde{\mathbf{z}})^T H(\tilde{\mathbf{z}})(\mathbf{z}-\tilde{\mathbf{z}})$ for $h(\mathbf{z})$ in $P(\mathcal{D}|\mathcal{M}_j) = \int e^{h(\mathbf{z})}d\mathbf{h}(\mathbf{z})$. However, finding the maximum likelihood point $\tilde{\mathbf{z}}$ usually requires expensive numerical computations [180]. Accordingly, further approximations of $P(\mathcal{D}|\mathcal{M}_j)$ can be achieved by replacing $\tilde{\mathbf{z}}$ with $\mathbf{z}^*$ the maximum likelihood estimate of $h^*(\mathbf{z}) = \log\big(f(\mathcal{D}|\mathbf{z},\mathcal{M}_j)\big)$, and $H(\tilde{\mathbf{z}})$ with $H^*(\mathbf{z}^*)$, minus the inverse Hessian of the log likelihood or Fisher's information matrix [180, 30]. Schwarz gave the BIC approximation for (5.10) when the number of observations are large [163];

$$\log\big(P(\mathcal{D}|\mathcal{M}_j)\big) \approx \log\big(f(\mathcal{D}|\mathbf{z}^*,\mathcal{M}_j)\big) - \frac{d}{2}\log(N) \tag{5.11}$$

where N is the number of observations. Raftery et al. [141] demonstrated the successful demonstration of Schwarz's approximation in a survival analysis problem. However, McCulloch et al. [111] showed that the BIC approximation may result in poor estimation when the number of observations are small. Since there is explicitly no prior probability distribution in Schwarz's approximation, it is formally not a Bayesian selection, but it may be implicitly considered as a Bayesian under a "unit information prior" [30, 89] or a "normalized Jeffreys prior" [30, 189].

Several variants of the Laplace's approximation have been proposed in the Bayesian literature using linear regression of $h(\mathbf{z})$. In particular, shrinkage estimators (e.g. ridge regression) are very popular [73]. In the family of shrinkage estimators, the Lasso method [179] is the most prominent algorithm which minimizes the least square error of the regression with an upper bound on the uncertain parameters. However, for the reason that the Laplace's approximation is basically a linear estimation around the

maximum likelihood point, these methods are not accurate for the cases where $h(\mathbf{z})$ is considerably nonlinear.

For nonlinear models, the efficiency of MCMC to explore the posterior distribution is greatly enhanced when rapidly computable closed form expressions for the marginal likelihoods $f(\mathcal{D}|\mathbf{z}, \mathcal{M}_j)$ are available. In this work, as it will be thoroughly explained in the next section, we utilize the polynomial chaos to achieve an analytical expression for the likelihood and compute the multidimensional integral of the equation (5.6) by substituting the polynomial chaos proxy for reservoir simulator. As we assume under each model class the prior distributions $P(\mathbf{z}|\mathcal{M}_i)$ are all multivariate Gaussian distributions, the polynomial chaos expansion gives an exponential convergence rate for the estimation of likelihood distribution.

If the object of Bayesian model selection is not only to identify the most probable model but to predict future observations as well, it is best to consider the panoply of models and the inferences or predictions they would give [41, 141]. A formal Bayesian solution to the predictive Bayesian model selection, was first proposed by Leamer [98]. The posterior probability of the mixture distribution $f(\mathbf{z}|\mathcal{D})$ can be expressed as

$$f(\mathbf{z}|\mathcal{D}) = \sum_{i=1}^{K} f(\mathbf{z}|\mathcal{D}, \mathcal{M}_i) P(\mathcal{M}_i|\mathcal{D}). \tag{5.12}$$

Consequently, if Δ is the quantity of interest, the posterior mean of Δ is given by

$$E(\Delta|\mathcal{D}) = \sum_{i=1}^{K} E(\Delta|\mathcal{D}, \mathcal{M}_i) P(\mathcal{M}_i|\mathcal{D}). \tag{5.13}$$

Madigan and Raftery [106] demonstrated that averaging over all the models in this fashion provides more accurate predictions. Raftery, Madigan and Hoeting [141] called this method the Bayesian Model Averaging (BMA). Several authors have studied the advantages of BMA and the costs of ignoring model uncertainty [61, 41, 141, 29, 89].

By applying the Bayes's rule, we can expand (5.12) as

$$\begin{aligned} f(\mathbf{z}|\mathcal{D}) &= \frac{f(\mathbf{z})}{P(\mathcal{D})} \cdot f(\mathcal{D}|\mathbf{z}) = \frac{f(\mathbf{z})}{P(\mathcal{D})} \Big(\sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) f(\mathcal{M}_i|\mathbf{z}) \Big) \\ &= \frac{f(\mathbf{z})}{f(\mathcal{D})} \Big(\sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) \frac{f(\mathbf{z}|\mathcal{M}_i) P(\mathcal{M}_i)}{f(\mathbf{z})} \Big) \\ &= \sum_{i=1}^{K} f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i) \frac{f(\mathbf{z}|\mathcal{M}_i) P(\mathcal{M}_i)}{P(\mathcal{D})}. \end{aligned} \tag{5.14}$$

where $f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i)$ is the likelihood of $\mathcal{D}$ under each cluster $\mathcal{M}_i$.

Example 5.3.1. ***Bayesian model selection for the mixture distribution of two Gaussian***

Let z be a random variable whose probability distribution function can be expressed as a mixture of two Gaussian probability distribution;

$$f(z) = \frac{1}{2}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-2)^2}{2}}\Big) + \frac{1}{2}\Big(\frac{1}{\sqrt{2\pi}}e^{-\frac{(z+2)^2}{2}}\Big). \tag{5.15}$$

Consequently, z can be effectively clustered into two different model, $\mathcal{M}_1$ and $\mathcal{M}_2$ where $P(\mathcal{M}_1) = P(\mathcal{M}_2) = \frac{1}{2}$ and

$$\begin{aligned} f(z|\mathcal{M}_1) &: \mathcal{N}(2,1) \\ f(z|\mathcal{M}_2) &: \mathcal{N}(-2,1). \end{aligned} \tag{5.16}$$

Assume the underlying physical system is $G(z) = e^z$ and the observed data $\mathcal{D} = 2$. For the likelihood distribution we consider the simple misfit formulation of Chapter 4:

$$f(\mathcal{D}|z) = ke^{-\frac{(G(z)-D)^2}{\sigma^2}} = ke^{-\frac{(e^z-D)^2}{\sigma^2}}. \tag{5.17}$$

Let the variance of the error in observation be $\sigma^2 = 1$. We want to calculate the

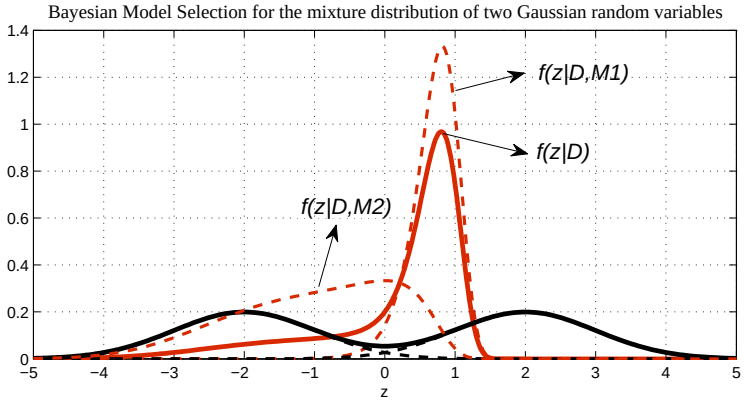

Figure 5.2: The red line shows the posterior probability of the mixture distribution of Example 5.3.1. The posterior probability under each cluster is shown with the red-dotted lines.

posterior probability of each cluster, Bayes factor and the posterior probability of the mixture distribution given the observed data $\mathcal{D}$. From the equation (5.8) we obtain

$$\begin{aligned} \frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} &= \frac{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_1)f(\mathbf{z}|\mathcal{M}_1)dz}{\int f(\mathcal{D}|\mathbf{z},\mathcal{M}_2)f(\mathbf{z}|\mathcal{M}_2)dz} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)} \\ &= \frac{\int e^{-\frac{(e^z-2)^2}{\sigma^2}}e^{-\frac{(z-2)^2}{2}}d\mathbf{z}}{\int e^{-\frac{(e^z-2)^2}{\sigma^2}}e^{-\frac{(z+2)^2}{2}}d\mathbf{z}} = 2.327. \end{aligned} \tag{5.18}$$

We also know the fact that $P(\mathcal{M}_1|\mathcal{D}) + P(\mathcal{M}_2|\mathcal{D}) = 1$*, then we achieve*

$$\begin{aligned} P(\mathcal{M}_1|\mathcal{D}) &\simeq 0.6994 \\ P(\mathcal{M}_2|\mathcal{D}) &\simeq 0.3006. \end{aligned} \tag{5.19}$$

The mixture posterior probability distribution of z *can also be computed via (5.14)*

$$f(z|\mathcal{D}) = const \cdot \big(\frac{1}{2} \cdot e^{-\frac{(e^z-2)^2}{\sigma^2}} e^{-\frac{(z-2)^2}{2}} + \frac{1}{2} \cdot e^{-\frac{(e^z-2)^2}{\sigma^2}} e^{-\frac{(z+2)^2}{2}}\big). \tag{5.20}$$

Figure 5.2 illustrates the prior and the posterior probability of z *under each cluster, in addition to the prior and the posterior probability of the mixture distribution.*

5.4 Bayesian Inference Using The Cluster-based Polynomial Chaos Proxy

In Chapter 4 we proposed an efficient method of uncertainty quantification for complex geological structures. We used the polynomial chaos expansion along with Markov Chain Monte Carlo method to probe posterior distribution of uncertain parameters given observed data. For complex geological structures where the number of random variables are large and their dependence are usually highly nonlinear, we applied nonlinear dimensionality reduction techniques to efficiently represent a random field in terms of a smaller number of random variables. We also discussed that for the non-Gaussian random variables the convergence rate of the standard Hermite polynomial chaos expansion is slow. It was demonstrated that the estimation quality of the posterior distribution depends on the order of the polynomial chaos proxy. However, more trial runs are required to achieve higher-order polynomial chaos expansions, which is computationally expensive. To resolve the issue, for a given number of trial runs N, we implemented the generalized polynomial chaos where the orthogonal basis are constructed adaptively to the input distribution, to achieve the optimum convergence rate.

However, for the practical problems we are provided with a set of realizations rather than a specific probability distribution and we need to use the numerical methods to estimate the density and construct polynomial chaos basis. One special case is when the realizations can be distinctly clustered into K different clusters where under each cluster the probability distribution can be reasonably approximated by Gaussian distribution. In general, all probability density functions can be expressed in terms of weighted Gaussian probability density functions, centered at each of the data points and then taking the average to yield Gaussian kernel density estimation [44]. Hence, the model-based clustering approach can be extended to problems where the number of clusters

K is unknown and can be optimized by study the distribution of each cluster. In this work, without loss of generality, we consider geological structures where the number of clusters K is pre-determined by geological intuition.

The objective is first to find an analytical expression for the posterior probability distribution of each cluster (and the mixture) given observed data, and second to compare different models (clusters) by the Bayesian model selection discussed in Section 5.3.

To estimate the posterior probability of each clusters, we follow the same routine proposed in Chapter 4, by using the standard polynomial chaos expansion as a proxy substitute of the actual reservoir simulator. Having an analytical expression for the posterior probability distribution of each cluster, and consequently the mixture, we apply the Markov Chain Monte Carlo method to sample from the posterior distribution. Since running the proxy is cheap, the MCMC method can efficiently probe the posterior probability distribution of each cluster and the overall mixture.

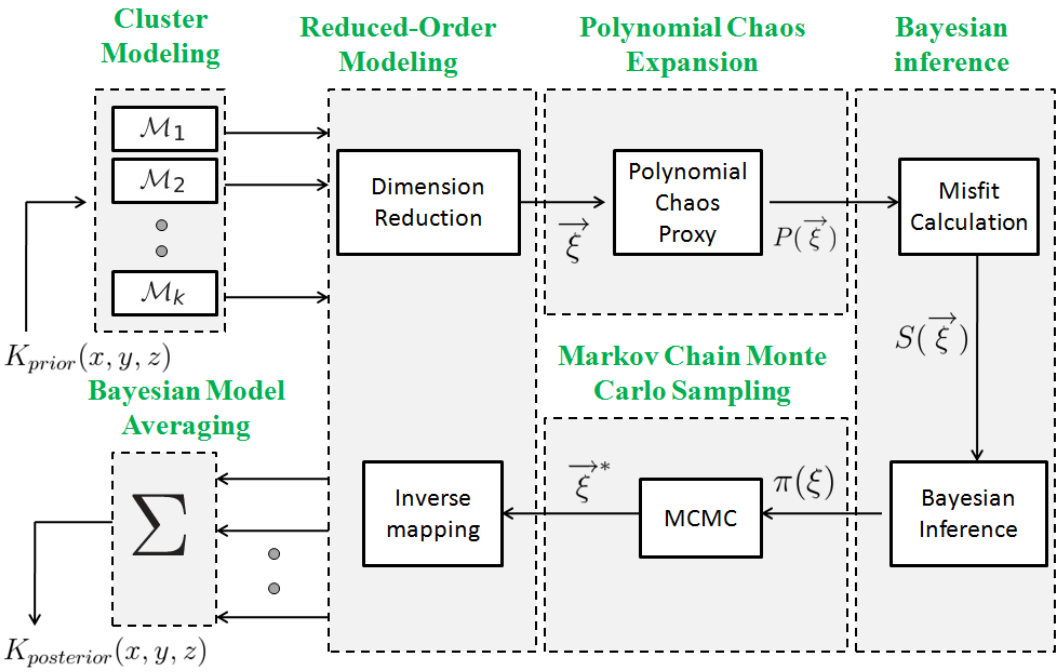

Figure 5.3: The framework for the Bayesian inference using the cluster-based polynomial chaos proxy

Figure 5.3 illustrates the framework for the Bayesian inference using the cluster-based polynomial chaos proxy. It consists of five main blocks; clustering component, dimension reduction component, the polynomial chaos proxy, the Bayesian model selection and the Markov Chain Monte Carlo sampling method. In the following we recap each component briefly as they have been discussed in detail in the previous chapters:

- **Clustering**

  Given a number of realizations N, we assume they can be effectively clustered into K different populations. We assume the probability distribution of each cluster is multivariate Gaussian distribution. If the realizations can not readily be clustered into different scenarios by the geological intuition, we use the statistical clustering tools such as Multi Dimensional Scaling (MDS) [40] or $K-$th mean clustering algorithm [102] to partition the realizations into K different subsets.

- **Dimensionality Reduction**

  In Chapter 3 we discussed that the efficiency of the polynomial chaos proxy greatly depends on the size of the problem, it is necessary to avoid the "curse of dimensionality". We explained the Karhunen-Loeve expansion as a linear dimensionality reduction tool and the kernel PCA as a nonlinear dimensionality reduction tool to efficiently break the size of a problem into reasonable number of random of variables which carry most information of a random field. As we assume the Gaussian distribution for each cluster, the Karhunen-Loeve expansion is the most effective tool to reduce the dimension while preserving the covariance matrix of each population.

- **Polynomial Chaos Proxy**

  In Chapter 2, we presented the constructions of the standard Wiener polynomial chaos expansions of random variables with finite second moments, and we showed that under the condition of the Cameron and Martin theorem, the polynomial chaos expansion guarantees the convergence in probability as the order of the expansion increases. The superior benefit of the polynomial chaos proxy as opposed to other surrogate models is that it guarantees convergence in probability as the order of the polynomial chaos expansion is increased. This implies that it does not suffer from the over-fitting issue that burden many other proxies. Given a limited number of trial runs, we described the intrusive and non-intrusive methods to estimate the polynomial chaos coefficients. In this work, we use the reduced-terms polynomial chaos proxy and apply regression-based PCM (2.4.2) to calculate the coefficients of the polynomial chaos proxy.

  We assume the probability distribution of $\mathbf{z}(\xi)$ under each clusters to be $P(\mathbf{z}|\mathcal{M}_j) : \mathcal{N}(\mu_j, \Sigma_j)$. Based on the given trial runs for each cluster, the polynomial chaos proxy $PC_{\mathcal{M}_j}$ for the cluster $\mathcal{M}_j$ is constructed via regression-based PCM. Regarding the exponential convergence rate of the standard polynomial chaos for

multivariate Gaussian random variables in each cluster $\mathcal{N}(\mu_j, \Sigma_j)$, even low order polynomial chaos suffices to give a reasonably accurate approximation.

Following the discussion for the misfit formulation given the observed data $\mathcal{D}$ in Section 4.2, under each cluster $\mathcal{M}_j$ the misfit surface approximation can be obtained by replacing the actual model with its corresponding polynomial chaos proxy substitute $PC_{\mathcal{M}_j}$;

$$\begin{aligned}\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{j}) &= \sum_i -\frac{(\mathbf{G}(\mathbf{z}) - \mathbf{d}^\mathbf{i}_\mathbf{obs})^2}{\mathbf{2}\sigma^\mathbf{2}_\mathbf{i}} \simeq \sum_i -\frac{(\mathbf{PC}_{\mathcal{M}_\mathbf{j}}(\mathbf{z}) - \mathbf{d}^\mathbf{i}_\mathbf{obs})^2}{\mathbf{2}\sigma^\mathbf{2}_\mathbf{i}} \\ &= S_0^{\mathcal{M}_j} + \sum_{i=1}^{N} S_{i,1}^{\mathcal{M}_j}\xi_i + \sum_{i=1}^{N} S_{ii,2}^{\mathcal{M}_j}(\xi_i^2 - 1) + \sum_{i=1}^{N-1}\sum_{k>i}^{N} S_{ik,2}^{\mathcal{M}_j}\xi_i\xi_k + \dots\end{aligned} \tag{5.21}$$

- **Bayesian Model Selection**

  The likelihood of the observed data under each cluster can be expressed as

$$f(\mathcal{D}|\mathbf{z}(\xi), \mathcal{M}_j) = ke^{-\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{j})}, \tag{5.22}$$

  where the misfit surface under each cluster $\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{j})$ can be computed via (5.21). Therefore, the Bayes factor of the equation (5.8) can be written as

$$\begin{aligned}\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} &= \frac{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_1) f(\mathbf{z}|\mathcal{M}_1) d\mathbf{z}}{\int f(\mathcal{D}|\mathbf{z}, \mathcal{M}_2) f(\mathbf{z}|\mathcal{M}_2) d\mathbf{z}} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)} \\ &\simeq \frac{\int e^{-\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{1})} f(\mathbf{z}(\xi)|\mathcal{M}_1) d\xi}{\int e^{-\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{2})} f(\mathbf{z}(\xi)|\mathcal{M}_2) d\xi} \times \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}.\end{aligned} \tag{5.23}$$

  Since we obtained an analytical expression for $\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{j})$ and $f(\mathbf{z}(\xi)|\mathcal{M}_j)$: $\mathcal{N}(\mu_j, \Sigma_j)$, the above integration can be efficiently calculated using the numerical techniques. Accordingly, we can achieve an analytical expression for the posterior probability of the mixture distribution by (5.14):

$$f(\mathbf{z}|\mathcal{D}) \simeq k\sum_{i=1}^{K} e^{-\mathbf{S}(\mathbf{z}(\xi)|\mathcal{M}_\mathbf{j})} f(\mathbf{z}(\xi)|\mathcal{M}_i) P(\mathcal{M}_i) \tag{5.24}$$

  The above formulation for the posterior mixture distribution is also called "Bayesian Model Averaging" [141].

- **Markov Chain Monte Carlo**

  Having an analytical expression for the posterior probability density of each cluster, we apply the Markov Chain Monte Carlo algorithm introduced in Section 4.3

to sample from the posterior distribution. We can either sample from the posterior probability of the mixture distribution (5.24) or sample from the posterior distribution under each cluster $f(\mathbf{z}(\xi)|\mathcal{D}, \mathcal{M}_j)$ and combine samples with respect to their corresponding Bayesian factor (5.8). The convergence of the Markov Chains under each cluster can be checked by the statistical diagnostic tools introduced in Section 4.3.4.

Clustering the uncertain domain $\mathbf{z}$ into K Gaussian clusters gives a better approximation of the posterior distribution comparing to the non-cluster approach proposed in Chapter 4, given N number of trial runs. Because the standard polynomial chaos gives the optimal exponential convergence rate when the input distributions are multivariate Gaussian. Hence, the polynomial chaos proxy under each cluster $PC_{\mathcal{M}_i}$ converges with an exponential rate to the true posterior distribution and consequently the mixture of posterior approximations converges to the true overall posterior distribution with an exponential rate. However, the convergence rate of the standard polynomial chaos proxy for the mixture of Gaussian distributions in the non-cluster approach is slow. In the following example we study the quality of the posterior estimation with the cluster and the non-cluster based approach.

Example 5.4.1. ***Bayesian inference using the cluster-based polynomial chaos proxy***

Consider the mixture of two Gaussian clusters of Example 5.3.1, $\mathcal{N}(-2,1)$ *and* $\mathcal{N}(2,1)$*, where the underlying physical system is* $G(z) = e^z$ *and the observed data* $\mathcal{D} = 2$*. Under each cluster* $\mathbf{z}$ *is Gaussian and can be expressed as:*

$$\mathbf{z} = \begin{cases} \xi + 2 & \mathbf{z} \in \mathcal{M}_1 \\ \xi - 2 & \mathbf{z} \in \mathcal{M}_2 \end{cases} \tag{5.25}$$

where ξ *is a normal random variable* $\mathcal{N}(0,1)$*. In the Example 2.3.1 of Chapter 2, the polynomial chaos expansion of* $G(\xi) = e^{\xi}$ *was derived as*

$$PC(G(\xi)) = \sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi) = e^{\frac{1}{2}}(1 + \xi + \frac{(\xi^2 - 1)}{2!} + \frac{(\xi^3 - 3\xi)}{3!} + \frac{(\xi^4 - 6\xi^2 + 3)}{4!} + \ldots). \tag{5.26}$$

Hence the polynomial chaos proxy under each cluster can be deduced via (5.26)

$$PC(G(\mathbf{z})) = \begin{cases} e^2\Big(\sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi)\Big) & \mathbf{z} \in \mathcal{M}_1 \\ e^{-2}\Big(\sum_{i=1}^{n} \frac{e^{\frac{1}{2}}}{i!} H_i(\xi)\Big) & \mathbf{z} \in \mathcal{M}_2 \end{cases} \tag{5.27}$$

Accordingly, the polynomial chaos proxy under each cluster can be expanded as

$$PC(G(\mathbf{z})) = \begin{cases} e^{\frac{5}{2}}\left(1+(z-2)+\frac{\left((z-2)^2-1\right)}{2!}+\frac{\left((z-2)^3-3(z-2)\right)}{3!}+\dots\right) & \mathbf{z}\in\mathcal{M}_1 \\ e^{\frac{-3}{2}}\left(1+(z+2)+\frac{\left((z+2)^2-1\right)}{2!}+\frac{\left((z+2)^3-3(z+2)\right)}{3!}+\dots\right) & \mathbf{z}\in\mathcal{M}_2 \end{cases} \tag{5.28}$$

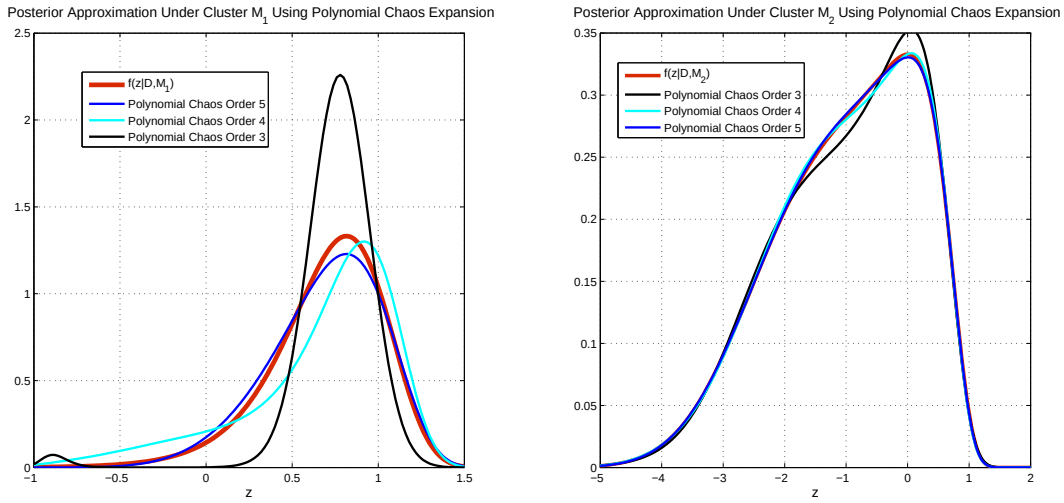

Figure 5.4: The posterior estimation under cluster $\mathcal{M}_1$ and $\mathcal{M}_2$ using the polynomial chaos expansion.

The likelihood under each cluster can be approximated by replacing $G(\mathbf{z})$ of the equation (5.17) with its corresponding polynomial chaos proxy,

$$\begin{aligned} f(\mathcal{D}|\mathbf{z},\mathcal{M}_1) &\simeq k\cdot e^{-\frac{(PC_{\mathcal{M}_1}(z)-D)^2}{\sigma^2}} = k\cdot e^{-\frac{(e^{\frac{5}{2}}(1+(z-2)+\frac{((z-2)^2-1)}{2!}+\dots-D)^2}{\sigma^2}} \\ f(\mathcal{D}|\mathbf{z},\mathcal{M}_2) &\simeq k\cdot e^{-\frac{(PC_{\mathcal{M}_2}(z)-D)^2}{\sigma^2}} = k\cdot e^{-\frac{(e^{\frac{-3}{2}}(1+(z+2)+\frac{((z+2)^2-1)}{2!}+\dots-D)^2}{\sigma^2}} \end{aligned} \tag{5.29}$$

Under each cluster, the Cameron-Martin Theorem 2.24 implies the exponential convergence rate for the polynomial chaos expansion to $G(\mathbf{z})$. Consequently, the posterior distribution under each cluster can be obtained by the equation (5.14). Figure 5.4 demonstrates the fast convergence rate of the estimation of the posterior distribution under each cluster using the polynomial chaos proxy to the exact answer derived in Example 5.3.1. The estimation of the posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy is shown on Figure 5.5.

However, in the non-cluster approach, the convergence rate of the standard polynomial chaos expansion is slow. Figure 5.6 illustrates the slow convergence of non-cluster

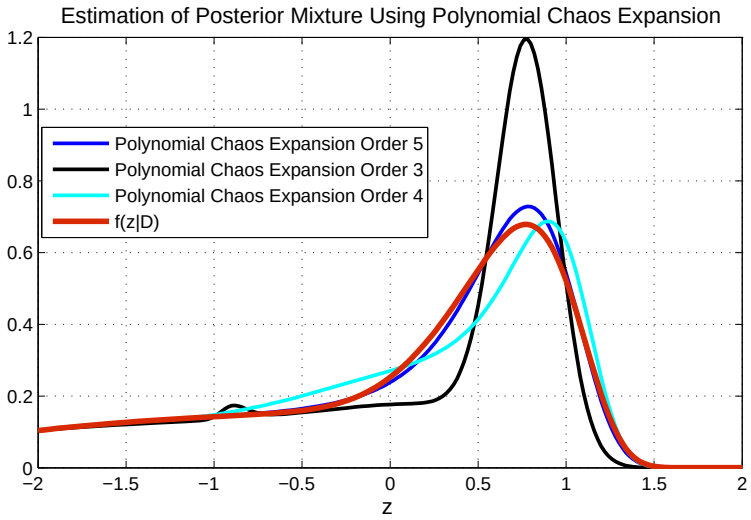

Figure 5.5: The estimation of the posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy.

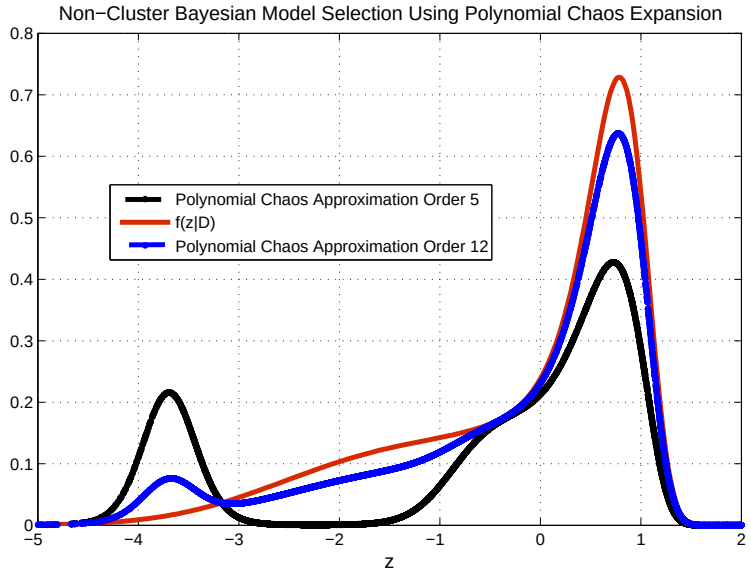

Figure 5.6: The estimation of the posterior probability of the mixture distribution using the standard non-cluster based polynomial chaos expansion.

based polynomial chaos proxy. As we discussed earlier in Example 2.3.2 of Chapter 2, the standard polynomial chaos expansion is the optimal choice for the Gaussian random variables but for the non-Gaussian random variables (e.g., the mixture of Gaussian) the higher order of polynomial chaos expansion may be required for the accurate approximation of the target distribution.

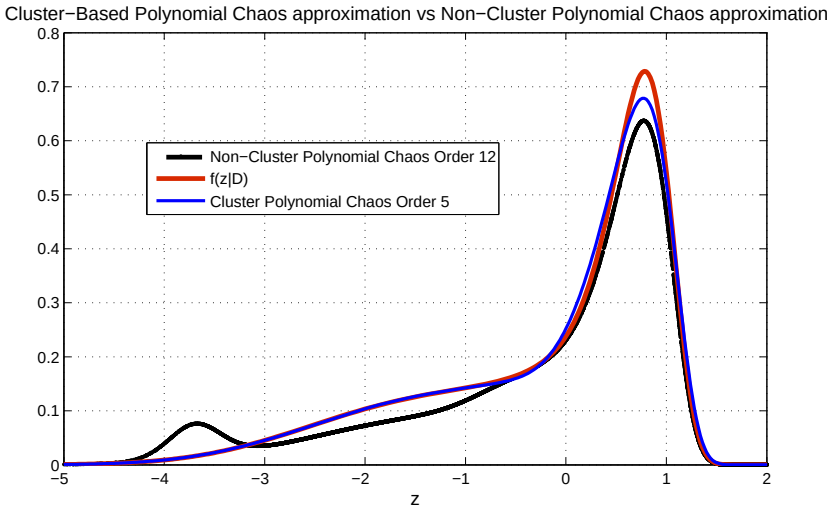

Figure 5.7: The cluster-based polynomial chaos proxy is more efficient that the non-cluster approach using the standard polynomial chaos expansion.

Figure 5.7 compares the accuracy in the estimation of the posterior probability of the mixture distribution using the cluster-based polynomial chaos proxy versus the standard non-cluster polynomial chaos proxy. It demonstrates that even the low order (4^{th} order) of the polynomial chaos expansion in the cluster-based approach gives a better approximation of posterior distribution than the polynomial chaos expansion of 12^{th} order in the non-cluster approach.

Accordingly, we can approximate the Bayes factor of the equation (5.18) by replacing $G(\mathbf{z})$ with the polynomial chaos proxy under each cluster. Table 5.1 compares the Bayes factor approximation by the polynomial chaos proxy of different order to the exact answer computed in Example 5.3.1. It is also instructive to compute the Bayes factor by the Laplace approximation and compare the accuracy with the polynomial chaos approximation. The Laplace's approximation under each cluster can be obtained by first computing the log of likelihood $h(\mathbf{z}|\mathcal{M}_i) = \log(f(\mathcal{D}|\mathbf{z}, \mathcal{M}_i)f(\mathbf{z}|\mathcal{M}_i)$ for each cluster;

$$\begin{aligned} h(\mathbf{z}|\mathcal{M}_1) &= \log(e^{-\frac{(e^{\mathbf{z}}-2)^2}{\sigma^2}} e^{-\frac{(\mathbf{z}-2)^2}{2}}) = -\frac{(\mathbf{z}-2)^2}{2} - \frac{(e^{\mathbf{z}}-2)^2}{\sigma^2} \\ h(\mathbf{z}|\mathcal{M}_1) &= \log(e^{-\frac{(e^{\mathbf{z}}-2)^2}{\sigma^2}} e^{-\frac{(\mathbf{z}+2)^2}{2}}) = -\frac{(\mathbf{z}+2)^2}{2} - \frac{(e^{\mathbf{z}}-2)^2}{\sigma^2}. \end{aligned} \tag{5.30}$$

Next we find the maximum likelihood point under each cluster by solving $h'(\mathbf{z}|\mathcal{M}_i) = 0$

where $\sigma^2 = 1$*;*

$$\begin{aligned} h(\mathbf{z}_1^*|\mathcal{M}_1) &= -(\mathbf{z}_1^* - 2) - \frac{2(e^{\mathbf{z}_1^*} - 2)e^{\mathbf{z}_1^*}}{\sigma^2} = 0 \Rightarrow \mathbf{z}_1^* = 0.8 \\ h(\mathbf{z}_2^*|\mathcal{M}_2) &= -(\mathbf{z}_2^* + 2) - \frac{2(e^{\mathbf{z}_2^*} - 2)e^{\mathbf{z}_2^*}}{\sigma^2} = 0 \Rightarrow \mathbf{z}_2^* = 0. \end{aligned} \tag{5.31}$$

The Laplace's method estimates $h(\mathbf{z}|\mathcal{M}_i)$ *by the second order Taylor expansion around the maximum likelihood point;* $h(\mathbf{z}|\mathcal{M}_i) \simeq h(\mathbf{z}^*|\mathcal{M}_i) + \frac{1}{2}h''(\mathbf{z}^*|\mathcal{M}_i)(\mathbf{z} - \mathbf{z}^*)^2$*, noting that* $h'(\mathbf{z}^*|\mathcal{M}_i) = 0$*;*

$$\begin{aligned} h(\mathbf{z}|\mathcal{M}_1) &\simeq -0.77086 - \frac{11.911}{2}(\mathbf{z} - 0.8)^2 \\ h(\mathbf{z}|\mathcal{M}_2) &\simeq -3 - \frac{1}{2}\mathbf{z}^2. \end{aligned} \tag{5.32}$$

Consequently, we can approximately compute the Bayes factor via the Laplace's method as

$$\frac{P(\mathcal{M}_1|\mathcal{D})}{P(\mathcal{M}_2|\mathcal{D})} = \frac{\int e^{h(\mathbf{z}|\mathcal{M}_1)}d\mathbf{z}}{\int e^{-0.77086-\frac{11.911}{2}(\mathbf{z}-0.8)^2}d\mathbf{z}} \simeq \frac{\int e^{h(\mathbf{z}|\mathcal{M}_1)}d\mathbf{z}}{\int e^{-3-\frac{1}{2}\mathbf{z}^2}d\mathbf{z}} \simeq 2.69167. \tag{5.33}$$

Table 5.1 shows that the polynomial chaos approximation is generally more accurate than the Laplace's method to compute the Bayes factor. This is mainly due to the fact that in approximating $G(\mathbf{z})$ *the polynomial chaos proxy takes the nonlinear terms into account while the Laplace methods uses a linear estimation around the maximum likelihood point.*

Table 5.1: Bayes factor estimation using the polynomial chaos proxy compared to the Laplace's method

	The Bayes Factor
The exact solution	2.327
The Laplace's method	2.691
Polynomial chaos proxy order 3	2.494
Polynomial chaos proxy order 4	2.383
Polynomial chaos proxy order 5	2.327

5.5 Case Study

Different interpretations of the sedimentary environments are one of the largest sources of uncertainty in reservoir modeling. Several plausible scenarios may be offered as reservoir interpretation based on seismic data, well log, etc. In the context of facies modeling, this could result in several descriptions of facies structures, associations, connectivity

and geometry. One quantitative approach to handle different geological scenarios is to provide alternative training images from which several reservoir models can be generated. The problem we address here is to quantitatively compare the plausibility of these training images and also sample from the posterior distribution within each scenario.

The pioneering work to integrate multiple reservoir descriptions using alternative training images was proposed by Caers and Scheidt (2008) [157]. They used a distance metric approach based on multi-dimensional scaling (MDS) to explore the space of the reservoir realizations. The method was further developed in [158] where MDS was used to compare different production responses of realizations obtained from different scenarios with the actual production data. Rojas [150] used a multi-class Support Vector Machine Classifier (SVM) [182] to differentiate between facies scenarios represented by multiple training images. Lately, Park et al. [133] proposed a Bayesian-based approach to calculate the posterior possibility of each scenario using the probability perturbation method. Here we use the cluster-based polynomial chaos proxy for Bayesian model selection. The case study demonstrates how the approach helps to quantify uncertainty between different training images.

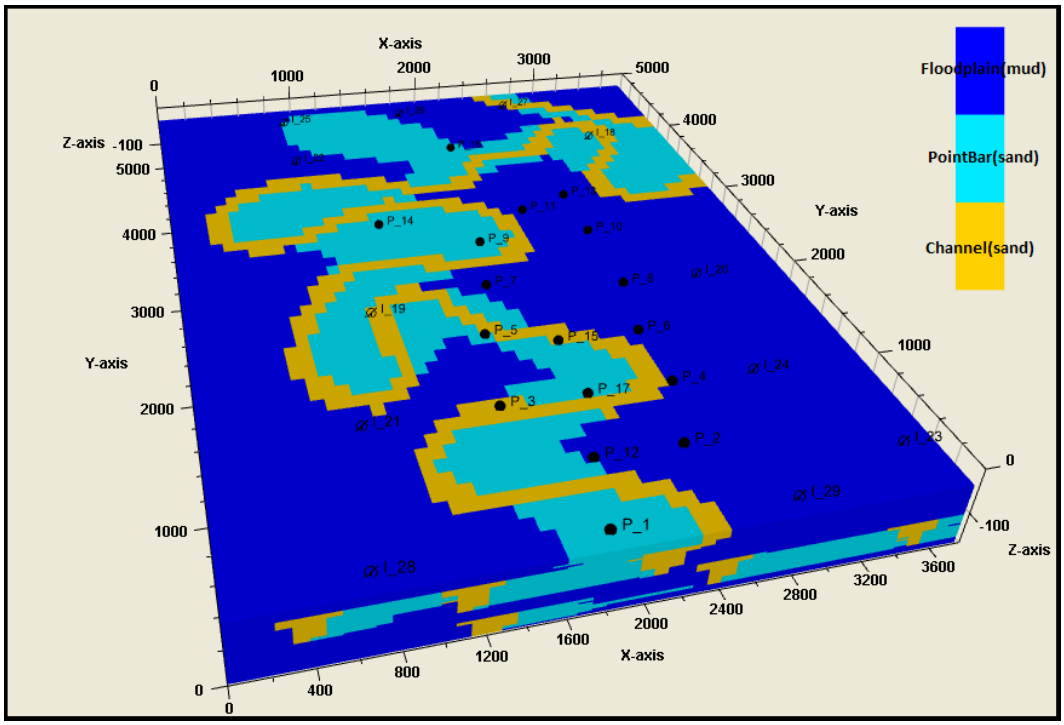

Figure 5.8: The reference case for the second layer of Stanford VI, with three facies: the floodplain, the point bar and the channel.

We study the plausibility of three training images based on different geological

interpretation of the second layer of synthetic Stanford VI reservoir ([28]). The reservoir is $3.75Km$ wide (East-West) and $5.0Km$ (North-South) long, with a shallowest top depth of 2.5 Km and deepest top depth of $2.7Km$. It consists of three layers with thickness of $80m$, $40m$ and $80m$. The Stanford VI reservoir is discretized into $150\times200\times200$ cells while the dimension of the grid cell is $25m$ in the x and y direction and $1m$ in the z direction. The stratigraphy of the Stanford VI reservoir shows a prograding fluvial channel system, where deltaic deposits represented in the third layer were deposited first and followed by meandering channels in the second layer and sinuous channel in the first layer. See [28] for the detailed description of the Stanford VI reservoir.

The second layer consists of meandering channel represented by four facies: the floodplain (shale deposits), the point bar (sand deposits that occur along the convex inner edges of the meanders of channels), the channel (sand deposits), and the boundary (shale deposits). Similar to Rojas [150], a simplified description of the second layer, which entails only three facies; the floodplain, the pointbar and the channel are modeled using three different training images, shown in Figure 5.9, Figure 5.10 and Figure 5.11. Given the fact that the prior probability of three scenarios are equal to each other $P(\mathcal{M}_1) = P(\mathcal{M}_2) = P(\mathcal{M}_3) = \frac{1}{3}$, we calculate the posterior probability of each scenario given the observed data $P(\mathcal{M}_1|\mathcal{D}), P(\mathcal{M}_2|\mathcal{D}), P(\mathcal{M}_3|\mathcal{D})$. The observed data includes the production profile of the second layer of the reservoir for 2000 days. The truth case used to generate the observation (the oil flow rate, the water cut and the pressure of all wells for 2000 days) is shown in Figure 5.8.

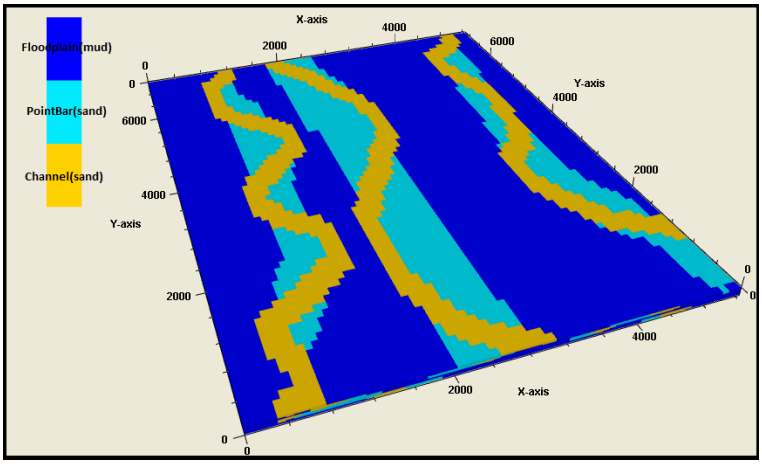

Figure 5.9: The training image of scenario $\mathcal{M}_1$

To calculate the posterior probability of each scenario, we generate $N = 500$ realizations for each training image using snesim [174]. We use the snesim algorithm introduced in Chapter 3 as a conditional simulation where the probability distribution

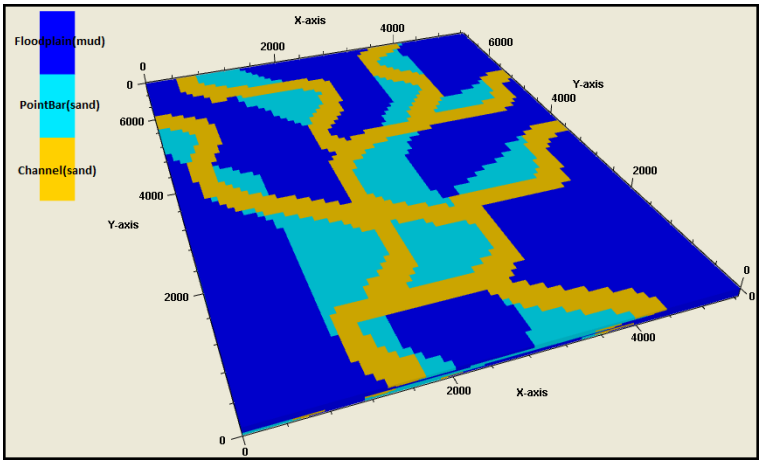

Figure 5.10: The training image of scenario $\mathcal{M}_2$

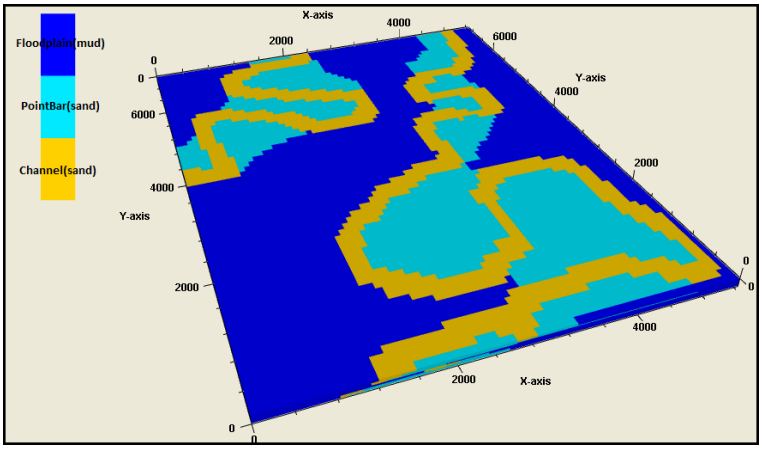

Figure 5.11: The training image of scenario $\mathcal{M}_3$

is sequentially generated from the training image and made conditional to data. All the facies realizations is then populated with three petrophysical properties: porosity, density and permeability as described in [28]. Hence, under each scenario we obtain 500 reservoir models. In Figure 5.13 we apply the principal component analysis to all 1500 facies realizations and plot the first two components. Three distinct clusters can be identified vividly and it demonstrates that the first two eigenvectors are enough to separate the realizations corresponding to the three training images. Figure 5.14 shows the Scree-graph of the realizations. As the realizations are well separated in three distinct clusters, then the differences between the groups will be picked up by the significant components of PCA.

Here we follow the procedure described in Section 5.4 which involves the following steps:

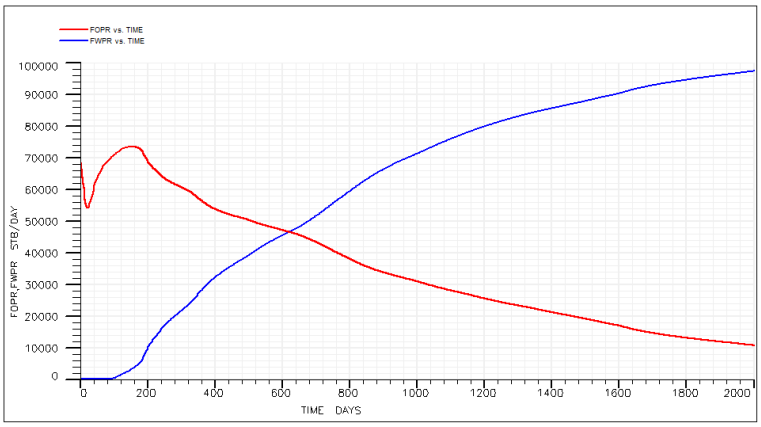

Figure 5.12: The cumulative oil (red) and water (blue) production rate of the second layer of Stanford VI for 2000 days.

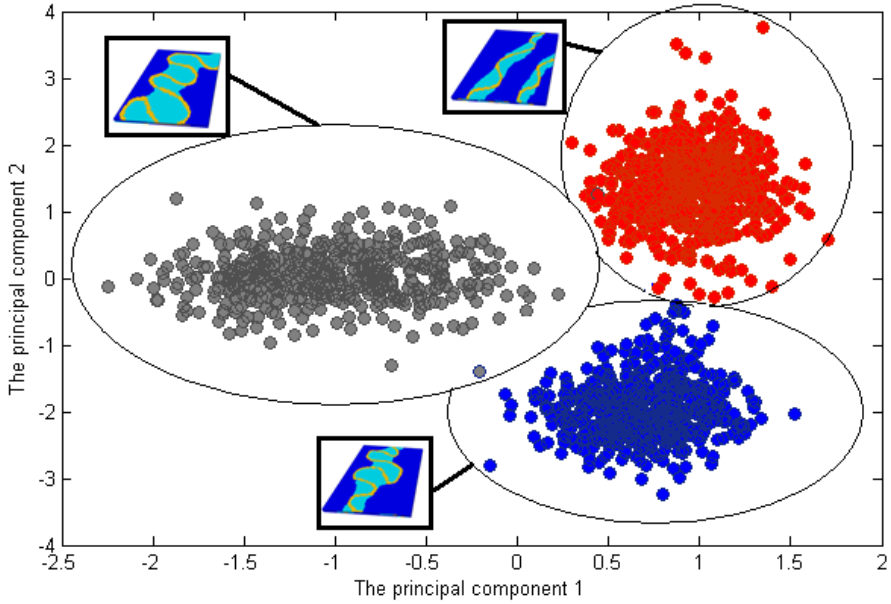

Figure 5.13: The principal component analysis for the realizations generated from three different training images.

- **Dimensionality Reduction**

  Under each cluster we apply the principal component analysis and discard the small eigenvalues. The within-groups covariance matrix describes the average variation of each group about its respective group mean. For this specific case, the within-groups covariance matrices of the three training images are similar to each other. Figure 5.15 illustrates the scree graph of the within-groups covariance matrix. For the Bayesian model selection, we preserve 10 eigenvectors under each

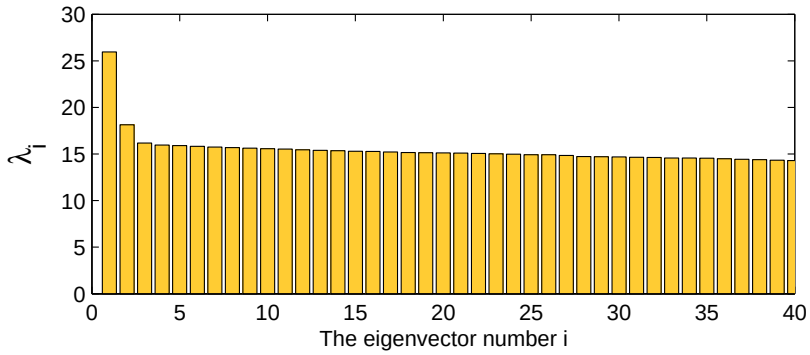

Figure 5.14: The scree graph of the 1500 realizations generated from three different training images.

cluster as the scree plots essentially flatten out thereafter. However, we compute the Bayes factor while preserving 2, 5, 10, 20 and 40 components.

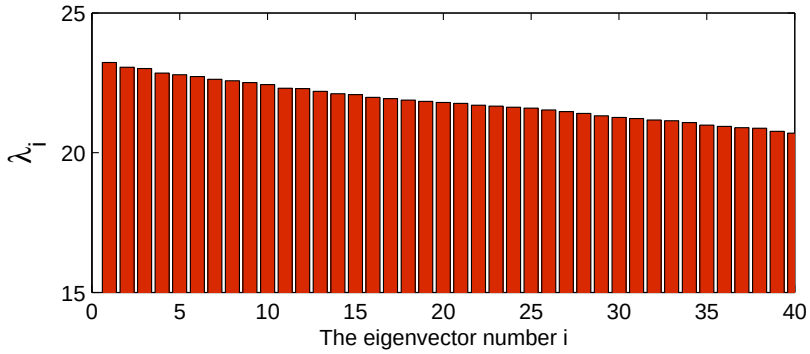

Figure 5.15: The scree graph of the realizations generated from the training image of scenario $\mathcal{M}_1$.

- **Polynomial Chaos Proxy**

  Using ECLIPSE for the flow simulation, we obtain the production profile of all the wells for the 1500 reservoir models generated earlier (trial runs). Consequently under each cluster, we apply the regression-based PCM using 1500 trial runs to compute the coefficients of the polynomial chaos representation for the oil and water flow rates at each well. For example, for the oil flow rate of the well P_1, we

obtain three different expressions for the polynomial chaos expansion as:

$$\begin{aligned}
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_1) &= Q_0^{op1|\mathcal{M}_1}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_1}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_1}(t)(\xi_i^2-1) + \dots \\
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_2) &= Q_0^{op1|\mathcal{M}_2}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_2}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_2}(t)(\xi_i^2-1) + \dots \\
\mathbf{Q^{op1}}(\xi,t|\mathcal{M}_3) &= Q_0^{op1|\mathcal{M}_3}(t) + \sum_{i=1}^{P} Q_{i,1}^{op1|\mathcal{M}_3}(t)\xi_i + \sum_{i=1}^{P} Q_{ii,2}^{op1|\mathcal{M}_3}(t)(\xi_i^2-1) + \dots
\end{aligned} \tag{5.34}$$

As far as the prior distribution under each cluster is assumed to be multivariate Gaussian, the Cameron-Martin Theorem 2.24 implies that the rate of convergence of the standard polynomial chaos is exponential. Hence, even a low order polynomial chaos will be reasonably accurate for the Bayes factor calculation. However, we examine the polynomial chaos expansion of order 2, 4, 6 and 8 respectively. For the 8^{th} order polynomial chaos expansion, to reduce the number of the terms, we use the impact factor introduced in Chapter 4 to discard the irrelevant terms.

Using the production data of the truth case (Figure 5.8) and the equation (5.21), we obtain the analytical expressions for the misfit surface under each cluster:

$$\begin{aligned}
\mathbf{S}(\xi|\mathcal{M_1}) &= S_0^{\mathcal{M}_1} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_1}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_1}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_1}(\xi_i\xi_j) + \dots \\
\mathbf{S}(\xi|\mathcal{M_2}) &= S_0^{\mathcal{M}_2} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_2}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_2}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_2}(\xi_i\xi_j) + \dots \\
\mathbf{S}(\xi|\mathcal{M_3}) &= S_0^{\mathcal{M}_3} + \sum_{i=1}^{P} S_{i,1}^{\mathcal{M}_3}\xi_i + \sum_{i=1}^{P} S_{ii,2}^{\mathcal{M}_3}(\xi_i^2-1) + \sum_{i=1}^{P-1}\sum_{j>i}^{P} S_{ij,2}^{\mathcal{M}_3}(\xi_i\xi_j) + \dots
\end{aligned} \tag{5.35}$$

We define the observational error $\sigma^2(t)$ similar to the previous chapter, i.e. $\sigma^2(t)$ for the oil flow rate of the well P_1 is defined as

$$\sigma(t) = \sigma_D\, Q_{obs}^{op1}(t) \tag{5.36}$$

where $Q_{obs}^{op1}(t) \in \mathcal{D}$ is the observed oil flow rate of P_1 (shown in Figure 5.16) and σ_D is the dimensionless observational error and assumed to be a constant for all the wells. We solve the problem for $\sigma_D^2 = 0.1, 0.15, 0.2$ and 0.25.

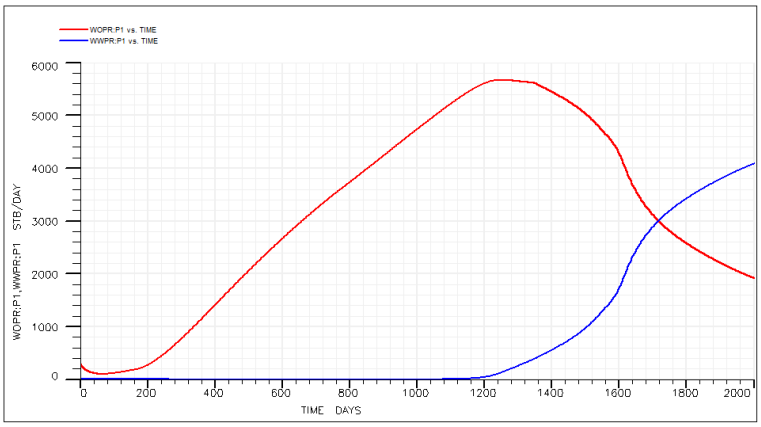

Figure 5.16: The oil (red) and water (blue) production rate of the well P1 for 2000 days. The observed data includes the production profile (the oil flow rate, the water cut and the pressure) of all the wells for 2000 days.

- **Markov Chain Monte Carlo Sampling**

  Using the equation (5.22), we obtain an analytical expression for the posterior distribution of the reduced dimension parameters ξ under each cluster, from which we can generate samples from the posterior reservoir models by the inversion map. Figure 5.17 shows samples from the posterior facies model under cluster $\mathcal{M}_1$ while 10 eigenvector is preserved and the order of the polynomial chaos proxy is 6. Figure 5.18 and 5.19 illustrates samples from the posterior facies models under cluster $\mathcal{M}_2$ and $\mathcal{M}_3$ respectively.

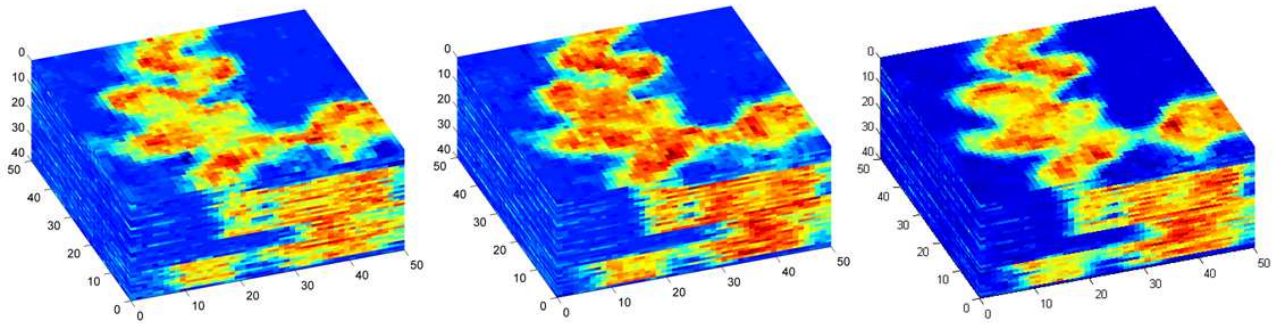

Figure 5.17: Samples from the posterior facies model under cluster $\mathcal{M}_1$.

- **Bayesian Model Selection**

  The high-dimensional integral of the equation (5.8) can be calculated using the numerical techniques developed in the literature [123, 167, 24, 124, 63]. Here we use the Quasi Monte Carlo [169] technique with the low discrepancy Sobol

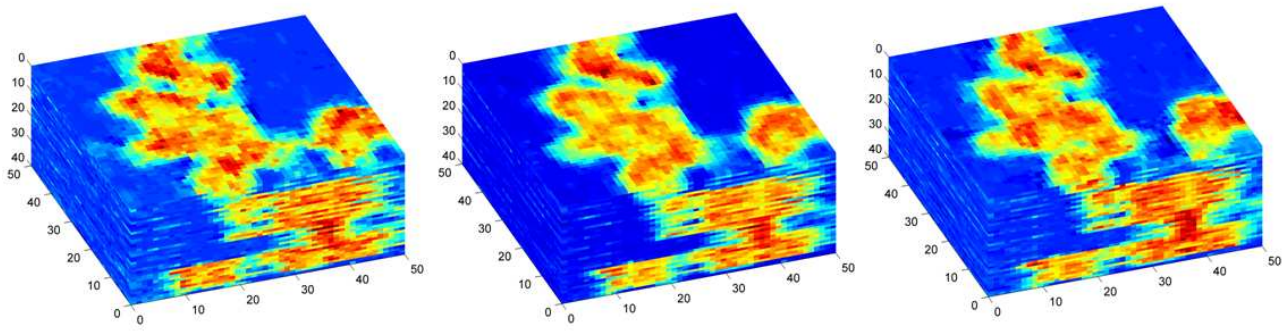

Figure 5.18: Samples from the posterior facies model under cluster $\mathcal{M}_2$.

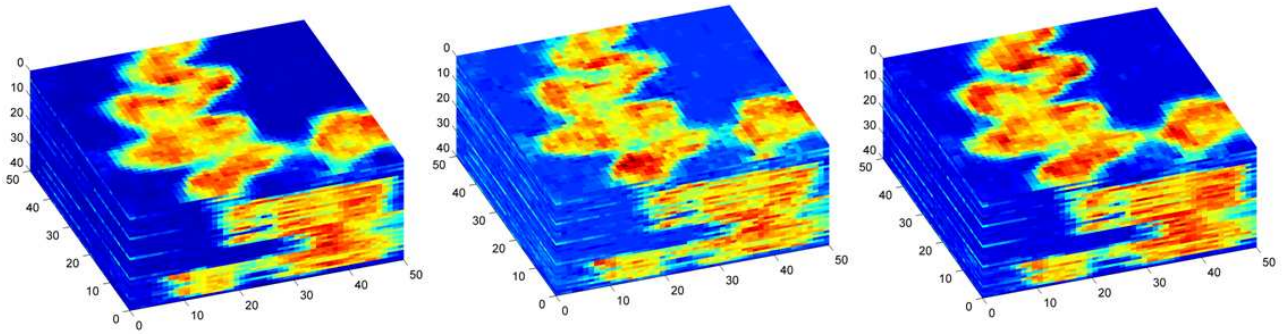

Figure 5.19: Samples from the posterior facies model under cluster $\mathcal{M}_3$.

sequence [123] whose rate of convergence is $\mathcal{O}(\frac{1}{N_s})$. N_s is the number of times that integrand has to be evaluated. Since running the polynomial chaos proxy is cheap, we can efficiently evaluate the integrand more than 10^9 times to achieve the desired accuracy.

The exact evaluation of the high-dimensional integral of (5.8) is not practically feasible. To examine the accuracy of the polynomial chaos proxy in the calculation of the Bayes factor, we retain 10 random variables of the reduced dimension space and perform a computationally extensive run of ECLIPSE for $N_s = 7880$ times. Subsequently we use the Quasi Monte Carlo method to calculate the integral numerically.

Figure 5.20 shows the posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$ using different orders of the polynomial chaos proxy. It is compared against the Quasi Monte Carlo method using the full reservoir simulator (ECLIPSE). It illustrates that the accuracy of

the approximation increases as the order of expansion increases. It also shows when the dimensionless observational error (σ_D) increases, the misfit formulation becomes more nonlinear and higher order polynomial chaos have to used for the accurate estimation of the posterior probability $P(\mathcal{M}_1|\mathcal{D})$. It also demonstrates that the model $\mathcal{M}_1$ (training image Figure 5.9) is the most probable model among others. The plausibility of $\mathcal{M}_1$ increases as we allow smaller error in the observation. For example, the posterior probability of $\mathcal{M}_1$ is $P(\mathcal{M}_1|\mathcal{D}) = 0.942$ when the dimensionless error in the observation is $\sigma_D^2 = 0.1$, while $P(\mathcal{M}_1|\mathcal{D})$ decreases to 0.67 when the observational error increases to $\sigma_D^2 = 0.2$.

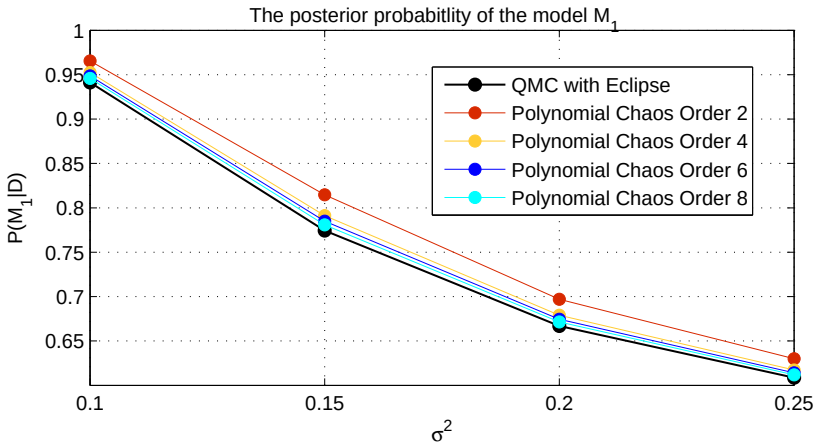

Figure 5.20: The posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

Figure 5.22 and Figure 5.21 illustrates the posterior probability of the model $\mathcal{M}_2$ and $\mathcal{M}_3$, $P(\mathcal{M}_2|\mathcal{D})$ and $P(\mathcal{M}_3|\mathcal{D})$ for different orders of the polynomial chaos expansion. It demonstrates that the model $\mathcal{M}_2$ is more plausible than the model $\mathcal{M}_3$. It also shows when the dimensionless observational error increases the plausibility of the model $\mathcal{M}_2$ and $\mathcal{M}_3$ increases.

Figure 5.23 shows the posterior probability of the model $\mathcal{M}_1$ when the number of retained eigenvalues increases. We used the polynomial chaos of order 6 for the proxy model. It demonstrates that the main difference between the probability of the models is picked up by the significant eigenvectors and the smaller eigenvectors do not contribute to the plausibility of the models remarkably. Accordingly, when the number of the retained eigenvectors increase from 10 to 40, $P(\mathcal{M}_1|\mathcal{D})$ does not change considerably. The dimensionless observational error for this experiment is assumed to be $\sigma_D^2 = 0.1$. Figure 5.24 shows the effect of increasing the number of eigenvectors on the posterior

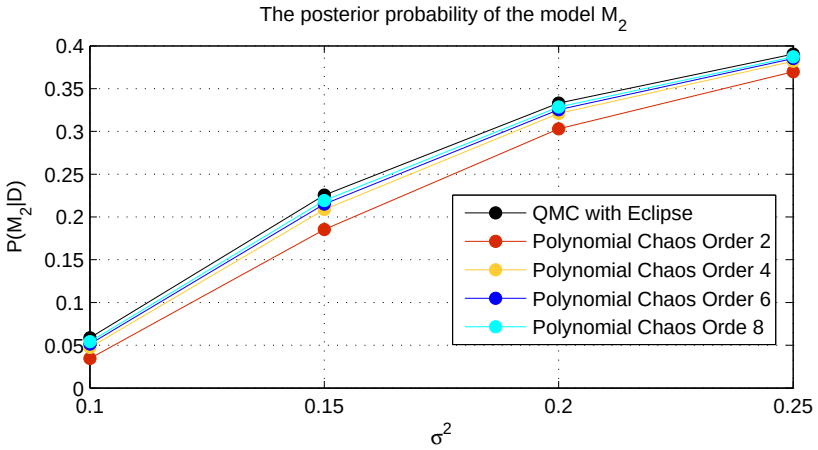

Figure 5.21: The posterior probability of the model $\mathcal{M}_2$, $P(\mathcal{M}_2|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

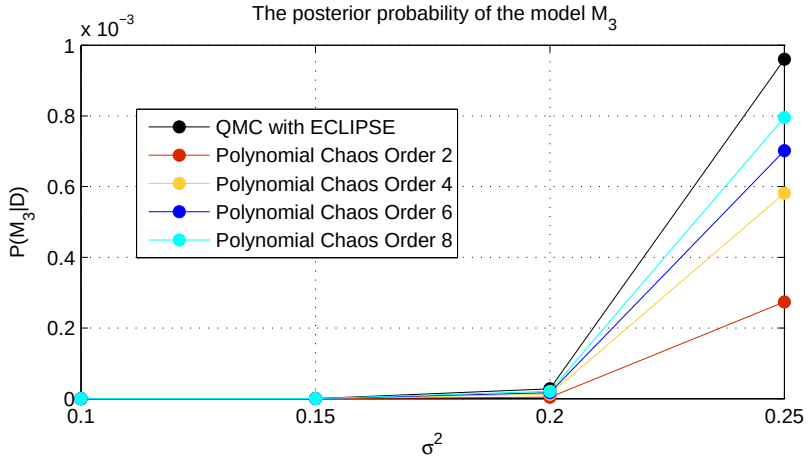

Figure 5.22: The posterior probability of the model $\mathcal{M}_3$, $P(\mathcal{M}_3|\mathcal{D})$, using different orders of the polynomial chaos proxy compared with the Quasi Monte Carlo method with ECLIPSE.

probability of the model $\mathcal{M}_2$.

Although in this specific case study the main difference between the plausibility of the models is picked up by the significant eigenvectors, the method proposed in this chapter can be coherently applied for the models that their difference delicately appears in small eigenvectors.

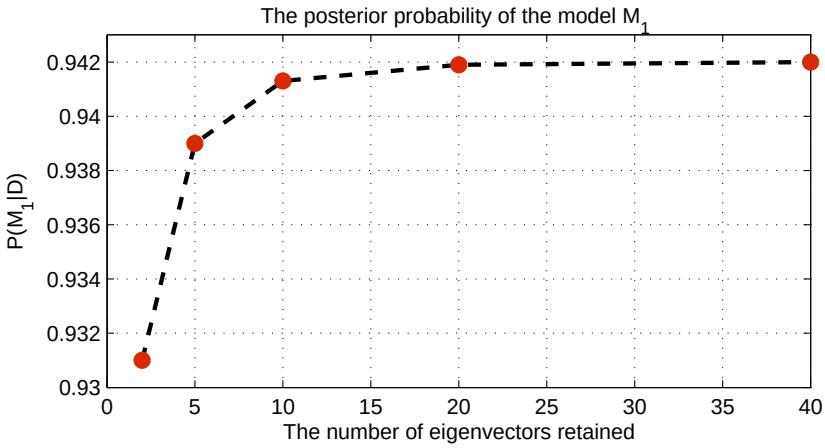

Figure 5.23: The posterior probability of the model $\mathcal{M}_1$, $P(\mathcal{M}_1|\mathcal{D})$, while different number of eigenvectors is retained.

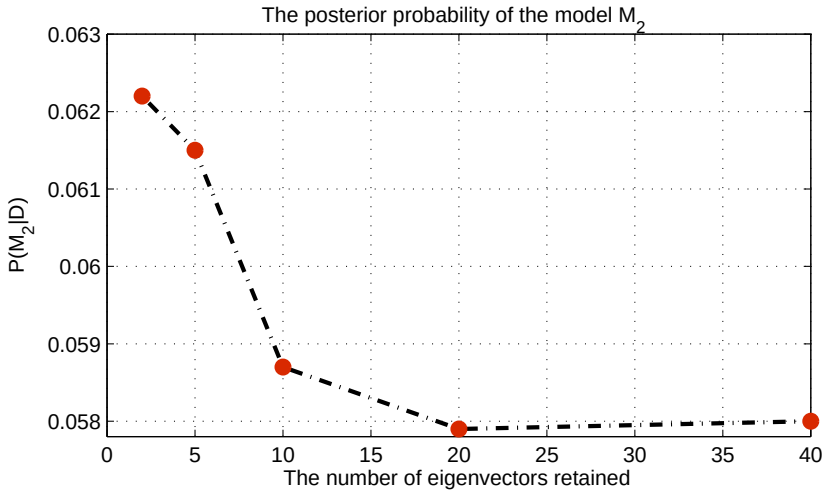

Figure 5.24: The posterior probability of the model $\mathcal{M}_2$, $P(\mathcal{M}_2|\mathcal{D})$, while different number of eigenvectors is retained.

5.6 Summary

In this chapter, we studied the "scenario uncertainty" where the prior model has a high level of discrete uncertainty. Such uncertainty usually comes from different interpretation of the sedimentary environments. For example, in real field applications, different scenarios of layering structure or fault interpretation may be plausible for a reservoir. The geological scenario usually has a considerable effect on flow response uncertainty. We use clustering approach to address the scenario uncertainty. Accordingly, we study the flow response uncertainty under each scenario and integrate the uncertainties to achieve the overall uncertainty. The objectives of clustering in our work are:

1. to make the problem less nonlinear under each scenario. Hence the polynomial chaos proxy converges faster for each cluster and the overall posterior distribution can be estimated more efficiently using different proxies.

2. to compare the possibility of different scenarios given data.

We apply a probabilistic approach for clustering, "Mixture modeling" where dataset are assumed to be rooted from a mixture of clusters with different probability density function. Then each individual sample does not exclusively belongs to a specific cluster and each cluster can contain a sample with a probability.

Our approach to cluster analysis in mixture modeling is based on the posterior probabilities of models, known as the "Bayesian Model Selection". In the sense that, we use Bayesian inference to calculate the posterior probability of each scenario and also the posterior probability density function of a parameter under each cluster.

For the posterior calculation, the Markov Chains Monte Carlo has been successfully practiced in the Bayesian framework. In reservoir modeling context, Gallagher et al. applied MCMC for Bayesian inference to select the optimal reservoir model [58]. However, in large problems MCMC may need more than 10^5 expensive runs of reservoir simulation and is not practically feasible. Our solution is to employ the polynomial chaos proxy to efficiently estimate posterior probability density function under each cluster and also calculate the posterior probability of each model. Consequently, we can compute the posterior distribution of a parameter for the mixture of the scenarios.

We applied the proposed cluster-based polynomial chaos proxy framework to study the plausibility of three training images based on different geological interpretation of the second layer of synthetic Stanford VI reservoir, based on the given data. We demonstrated that the proposed workflow can be efficiently used to calculate the posterior probability of each scenario and also sample from the posterior facies models within each scenario.

Chapter 6

Concluding Remarks

This dissertation addressed the application of the polynomial chaos expansions as a proxy model for history matching and uncertainty quantification in reservoir study. In this chapter, we present the concluding remarks and also the recommendations for the future research.

6.1 Summary and Conclusions

The key findings of this thesis are as follows:

- The application of the polynomial chaos expansions for history matching and uncertainty quantification is thoroughly studied in this work. The main advantage of the polynomial chaos proxy compared to other proxy models and response surfaces is that it is generally applicable and converges systematically all the time as the order of the expansion increases. In fact, the Cameron and Martin theorem states that the convergence of the standard polynomial chaos expansions is exponentially fast for Gaussian random variables. It also gives the error bound for the n^{th}-order polynomial chaos expansion and quantitatively illustrates how reliable the approximation is. For non-Gaussian random variables the convergence rate of the standard polynomial chaos might be slow. To improve the convergence rate for non-Gaussian random variables, the generalized polynomial chaos is presented that uses Askey-scheme to choose the optimal basis for the polynomial chaos expansions.

- One other advantage of approximating the misfit surface with the polynomial chaos expansions is that there are very efficient algorithms for polynomial optimizations. Semi-definite programming can efficiently compute the global minimum of strictly positive multivariate polynomials.

- For a given number of trial runs, the coefficients associated with the higher order polynomial chaos terms have larger estimation errors compared to the lower order terms, and therefore there is a balance between the number of trial runs and the maximum possible order of the polynomial chaos expansions that can be reliably used.

- For high-dimensional problems, the number of the polynomial chaos terms increases drastically as the order of the polynomial chaos expansions increases. Although different non-intrusive methods have been developed in the literature to address this issue, but still a large number of the simulation runs is required to compute high-order polynomial chaos expansions. The "sparsity of effects" principle implies that most models are mainly governed by the main-effects and only the lower-order interaction between the input random variables. Hence, for most of the practical applications the polynomial chaos expansions are considerably sparse and many terms could be neglected from the expansions. Accordingly, the number of the simulation runs required to achieve high-order polynomial chaos approximation of a forward model does not increase exceedingly as only the relevant terms are preserved in the reduced-terms polynomial chaos expansions. Several methods have been recently proposed to detect the sparsity pattern and achieve an accurate sparse polynomial chaos expansions, or adaptively construct the polynomial chaos basis to obtain a highly sparse polynomial chaos representation which has only few non-zero terms.

- In our proposed polynomial chaos proxy which integrates polynomial chaos expansion with the Karhunen-Loéve decomposition method (or kernel PCA to preserve multi-point statistics), the sparsity pattern of the polynomial chaos can be determined using the relative importance of K-L modes (or modes of kernel PCA in the feature space). In the sense that among the cross-terms associated with the high-frequency K-L modes (the modes related to the small eigenvalues) only few terms are relevant and the rest become adequately negligible. We introduced the impact factor for each term to specify the relevant terms in the polynomial chaos proxy and discard the rest. Accordingly, for high-dimensional problems, with a

reasonable number of simulation runs, the high-order reduced-terms polynomial chaos proxy can be efficiently constructed.

- Given N number of simulation runs that could be accommodated for a specific problem, and p number of retained eigenvectors, the maximum order of the polynomial chaos proxy that can be reliably used is determined by the following procedure:

  1. All the first order terms are retained, $\{\xi_1, \xi_2, \ldots, \xi_p\}$,
  2. The order of the polynomial chaos is increased to the next order, $PC_m(\xi, t)$,
  3. The insignificant terms in $PC_m(\xi, t)$ are dropped based on their corresponding impact factor and the pre-determined cut-off value ϵ,
  4. While the total number of the unknown coefficients in the reduced-terms polynomial chaos proxy is smaller than N, repeat the procedure.

  Consequently, the K-L modes appear with different orders in the reduced-terms polynomial chaos proxy. For high-dimensional problems, the proposed algorithm only preserves the pure terms of the high-frequency modes and as a result the curse of dimensionality is mitigated acutely.

- The use of the polynomial chaos proxy in the Bayesian context, where the polynomial chaos representation replaces the full forward model in the likelihood function, obtains an analytical expression for the reduced-dimension posterior distribution that can be evaluated orders of magnitude faster than the original posterior. Additionally, with the respect to the fact that running the proxy model is cheap, the Markov Chain Monte Carlo methods can be coherently used to efficiently sample from the reduced-dimension approximation of the posterior distribution.

- The relative importance of the K-L modes (or kernel PCA modes in the feature space) changes from the prior to the posterior, but still decays for higher frequency terms.

- Comparing to the traditional MCMC methods which typically needs more than 10^6 full reservoir simulation runs, the polynomial chaos proxy-based MCMC can probe the posterior distribution much faster, using a few hundreds of the full reservoir simulation runs. For remarkably nonlinear surface responses we demonstrated that the high-order polynomial chaos proxy can be used to preserve higher-order moments of the posterior distribution. The number of the simulation runs required to reasonably capture the nonlinear effects of the posterior distribution

using a high order polynomial chaos proxy is problem specific. In general, the accuracy of the polynomial chaos proxy depends on:

1. The quality of the input data (the trial simulation runs),
2. The order of the polynomial chaos expansions,
3. The number of eigenvectors retained.

For extremely nonlinear surface response, one can expect that the number of the simulation runs required to construct a reliable polynomial chaos proxy increases and the proposed method becomes less efficient. We discuss this issue in the next section as a future research recommendation.

- When the eigenvalues of the K-L decomposition are not reasonably fast decaying (large variance), the number of the relevant terms in the reduced-terms polynomial chaos representation increases and the number of the full reservoir simulation runs required to compute the polynomial chaos coefficients increases. Hence, the proposed method becomes less efficient when the random field can not be effectively expressed using the K-L decomposition or kernel PCA.

- The prior distribution for the reduced dimension random variables of the K-L decomposition is assumed to be uncorrelated multivariate Gaussian distribution. Consequently, the standard polynomial chaos will have an exponential convergence rate to the true misfit surface as the order of the expansions increases. The posterior distribution of the reduced-dimension random variables are not necessarily Gaussian and they might be even considerably correlated. For the non-Gaussian prior distributions, the generalized polynomial chaos proxy can be used to improve the convergence rate.

- When the prior model has a high level of discrete uncertainty, "scenario uncertainty", to improve the convergence rate for the polynomial chaos expansions, a mixture modeling approach is used for clustering to make the problem less nonlinear by considering each cluster separately. Hence, the polynomial chaos proxies converge faster for each cluster and can be integrated to estimate the overall posterior distribution more efficiently.

- A probabilistic framework based on the polynomial chaos proxy is proposed to study the plausibility of different geological interpretations of the sedimentary environments. To calculate the probability of each scenario, the Bayesian model selection is used to integrate over all possible parameters in each scenario with

respect to their corresponding priors and accordingly give a measure of how well a scenario is supported by observations. The main challenge in the practical implementation of the Bayesian model selection is that the calculation of the marginal likelihood for high-dimensional problems become infeasible. The polynomial chaos proxy provides an analytical expression for the posterior distribution under each scenario and allows for an efficient estimation of the marginal likelihood for each scenario.

- One quantitative approach to handle different geological scenarios is to provide alternative training images from which several reservoir models can be generated. We applied our cluster-based polynomial chaos proxy to integrate multiple reservoir descriptions using alternative training images and also sample from the posterior distribution within each scenario. Additionally, we quantitatively compared the plausibility of each training image with the Bayesian approach. In our case study, the main difference between the plausibility of the training images was picked up by the significant eigenvectors, but the proposed method can be coherently used to compare training images that their differences appear in smaller eigenvalues.

6.2 Recommendations for Future Work

- To find the global minimum of the misfit surface more accurately for history matching of highly nonlinear problems, the approximated global minimum found by the polynomial chaos proxy can be used as an initial guess for the gradient-based optimization algorithms. It will reduce the number of iterations required to obtain the extremum point, if the global minimum does not happen in highly improbable region.

- The idea of the reduced-terms polynomial chaos proxy can still be developed to obtain a highly sparse polynomial chaos representations by either adaptively constructing the polynomial chaos basis to generate a sparse pattern, or alternatively choosing the proper geological parameterization method. For physical parameters, most models are governed mainly by the pure terms and only the low-order cross-effects. Hence, the parameters can be chosen to carry more information about the physics of a system to increase the possibility of a sparse pattern.

- The accuracy and runtime speedup provided by the polynomial chaos proxy-based history matching are closely related to the number of the input random variables

(or the amount of energy retained by the eigenvectors from the random field) and the order of the polynomial chaos expansions. The study could be extended to evaluate the compromise between accuracy and speedup to achieve the optimum number of eigenvectors retained.

- For extremely nonlinear posterior distributions, where even a high-order polynomial chaos proxy is not accurate enough, or too many simulation runs required to obtain a reliable polynomial chaos proxy, the full MCMC methods seems to be the only option to probe the posterior distribution. However, we can increase the acceptance rate and make the Markov chain more efficient, by using the posterior distribution estimated by the polynomial chaos proxy as a proposal distribution for the MCMC method. This will increase the efficiency of the Markov chain and reduce the length of the chain to reach the equilibrium state.

- The polynomial chaos proxy can be coherently used with other variants of the MCMC methods. For example, the polynomial chaos proxy can be used in the Hamiltonian MCMC method to estimate the gradients of an objective function. It will mitigate the implementation of these algorithms and possibly improve their efficiencies.

- The polynomial chaos proxy can be used as an approximation model for a two-stage MCMC method, where at each step, it evaluates the likelihood of the point using the approximation model to decide if it is worth to run the full forward model for the proposed move. This procedure will fine-tune the posterior distribution estimated by the polynomial chaos proxy to account for the high-frequency effects.

References

[1] SI Aanonsen, G Naevdal, DS Oliver, AC Reynolds, and B Valles. The Ensemble Kalman Filter in reservoir engineering–a review. *SPE Journal*, 14(3):393–412, 2009. [cited at p. 81, 126]

[2] Hirotugu Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723, 1974. [cited at p. 136]

[3] Michael R Anderberg. Cluster analysis for applications. Technical report, DTIC Document, 1973. [cited at p. 134]

[4] C Andrieu, A Doucet, and CP Robert. Computational advances for and from Bayesian analysis. *Statistical science*, 19(1):118–127, 2004. [cited at p. 140]

[5] Christophe Andrieu, Nando De Freitas, and Arnaud Doucet. Sequential MCMC for Bayesian model selection. In *Higher-Order Statistics, 1999. Proceedings of the IEEE Signal Processing Workshop on*, pages 130–134. IEEE, 1999. [cited at p. 136]

[6] Christophe Andrieu and Arnaud Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *Signal Processing, IEEE Transactions on*, 47(10):2667–2676, 1999. [cited at p. 136]

[7] de Freitas N. Doucet A. Andrieu, C. and M. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003. [cited at p. 89, 90, 92, 94]

[8] Ivo Babuska, Raúl Tempone, and Georgios E Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004. [cited at p. 13, 63]

[9] Ivo Babuška, Raúl Tempone, and Georgios E Zouraris. Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Computer methods in applied mechanics and engineering*, 194(12):1251–1294, 2005. [cited at p. 13, 63]

[10] O Badru and CS Kabir. Well placement optimization in field development. In *SPE Annual Technical Conference and Exhibition*, 2003. [cited at p. 8]

[11] Mark Baker. Cluster computing white paper (version 2.0). *IEEE Task Force on Cluster Computing*, 2000. [cited at p. 6]

[12] John Barker, Cuypers Maarten, and Holden Lars. Quantifying uncertainty in production forecasts: Another look at the PUNQ-S3 problem. In *SPE Annual Technical Conference and Exhibition*, 2000. [cited at p. 8]

[13] Patrick Billingsley. *Convergence of probability measures*, volume 493. John Wiley & Sons, 2009. [cited at p. 15]

[14] John R Birge. Quasi-Monte Carlo approaches to option pricing. Technical report, 1995. [cited at p. 37]

[15] Géraud Blatman and Bruno Sudret. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique*, 336(6):518–523, 2008. [cited at p. 102]

[16] Géraud Blatman and Bruno Sudret. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25(2):183–197, 2010. [cited at p. 102]

[17] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, 2011. [cited at p. 102]

[18] Luciane Bonet-Cunha, Dean S Oliver, RA Redner, and AC Reynolds. A hybrid Markov chain Monte Carlo method for generating permeability fields conditioned to multiwell pressure data and prior information. *SPE Journal*, 3(3):261–271, 1998. [cited at p. 8]

[19] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. [cited at p. 45]

[20] DR Brouwer and JD Jansen. Dynamic optimization of water flooding with smart wells using optimal control theory. In *European Petroleum Conference*, 2002. [cited at p. 43, 67]

[21] WJ Browne and David Draper. Implementation and performance issues in the Bayesian and likelihood fitting of multilevel models. *Computational statistics*, 15:391–420, 2000. [cited at p. 94]

[22] Jef Caers. History matching under training-image-based geological model constraints. *SPE journal*, 8(3):218–226, 2003. [cited at p. 42]

[23] Jef Caers. *Petroleum geostatistics*. Richardson, TX: Society of Petroleum Engineers, 2005. [cited at p. 47, 48, 49, 50, 51]

[24] Russel E Caflisch. Monte carlo and quasi-monte carlo methods. *Acta numerica*, 1998:1–49, 1998. [cited at p. 158]

[25] Robert H Cameron and William T Martin. The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *The Annals of Mathematics*, 48(2):385–392, 1947. [cited at p. 12, 17, 20, 21, 28, 38, 133]

[26] Bradley P Carlin and Siddhartha Chib. Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 473–484, 1995. [cited at p. 136, 140]

[27] J. Carter and P. Ballester. A real parameter genetic algorithm for cluster identification in history matching. 30, 2004. [cited at p. 80, 81]

[28] S Castro, J Caers, and T Mukerji. The stanford vi reservoir. *Stanford Center for Reservoir Forecasting (SCRF), Annual Report*, 2005. [cited at p. 153, 154]

[29] Chris Chatfield. Model uncertainty. *Encyclopedia of Environmetrics*, 1995. [cited at p. 141]

[30] Hugh Chipman, Edward I George, Robert E McCulloch, M Clyde, Dean P Foster, and Robert A Stine. The practical implementation of Bayesian model selection. *Lecture Notes-Monograph Series*, pages 65–134, 2001. [cited at p. 136, 140]

[31] Merlise Clyde and Edward I George. Model uncertainty. *Statistical science*, pages 81–94, 2004. [cited at p. 136]

[32] Keith H Coats, LK Thomas, and RG Pierson. Compositional and black oil reservoir simulation. In *SPE Reservoir Simulation Symposium*, 1995. [cited at p. 5]

[33] ExxonMobil Corporation. Energy outlook to 2030, technical report from www.exxonmobil.com/energyoutlook. 2004. [cited at p. iii, 1, 2]

[34] Mary Kathryn Cowles and Bradley P Carlin. Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996. [cited at p. 94, 95]

[35] DR Cox. Tests of separate families of hypotheses, 1961. [cited at p. 136]

[36] Flavia Da Silva, Jose Rodrigues, Paulo Paraizo, Regis Romeu, Alvaro Peres, Rildo Oliveira, Iubatan Pinto, and Celio Maschio. Novel ways of parameterizing the history-matching problem. *SPE Journal-94875*, 2005. [cited at p. 48]

[37] Laurie P Dake. *Fundamentals of reservoir engineering*. Access Online via Elsevier, 1983. [cited at p. 1, 2]

[38] Henry Darcy. *Les fontaines publiques de la ville de Dijon: exposition et application*. Victor Dalmont, 1856. [cited at p. 3]

[39] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. *International Computer Science Institute, Technical Report*, pages 99–006, 1999. [cited at p. 57]

[40] Mark Leonard Davison. *Multidimensional scaling*. Wiley New York, 1983. [cited at p. 134, 145]

[41] David Draper. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 45–97, 1995. [cited at p. 141]

[42] Simon Duane, Anthony D Kennedy, Brian J Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987. [cited at p. 81]

[43] Joanna Dunkley, Martin Bucher, Pedro G Ferreira, Kavilan Moodley, and Constantinos Skordis. Fast and reliable Markov chain Monte Carlo technique for cosmological parameter estimation. *Monthly Notices of the Royal Astronomical Society*, 356(3):925–936, 2005. [cited at p. 95]

[44] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163, 2002. [cited at p. 143]

[45] AA Emerick and AC Reynolds. Combining the Ensemble Kalman Filter with Markov chain Monte Carlo for improved history matching and uncertainty characterization. *SPE J*, 17(2):418–440, 2012. [cited at p. 82]

[46] Bahar Erar. Mixture model cluster analysis under different covariance structures using information complexity. 2011. [cited at p. 134, 135, 136]

[47] Demet Erbas and Michael Christie. Effect of sampling strategies on prediction uncertainty estimation. In *Proceedings of the SPE Reservoir Simulation Symposium, SPE 106229, 26Ű28*, 2007. [cited at p. 81]

[48] Oliver G Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM Math. Model. Numer. Anal*, 46(2):317–339, 2012. [cited at p. 20]

[49] Geir Evensen. *Data assimilation: the ensemble Kalman filter*. Springer, 2009. [cited at p. 81, 126]

[50] Brian S Everitt. *The Cambridge dictionary of statistics in the medical sciences*. Cambridge University Press Cambridge, 1995. [cited at p. 134]

[51] Brian S Everitt and David J Hand. Finite mixture distributions. *Monographs on Applied Probability and Statistics, London: Chapman and Hall, 1981*, 1, 1981. [cited at p. 137]

[52] Richard E Ewing, Michael S Pilant, J Gordon Wade, and A Ted Watson. Estimating parameters in scientific computation: A survey of experience from oil and groundwater modeling. *IEEE Computational Science & Engineering*, 1(3):19–31, 1994. [cited at p. 7]

[53] Irving Fatt. The network model of porous media i. Capillary pressure characteristics. *Trans. AIME*, 207(7):144–159, 1956. [cited at p. 5]

[54] George S Fishman. *Monte Carlo: concepts, algorithms, and applications*, volume 1196. Springer New York, 1996. [cited at p. 8]

[55] FJT Floris, MD Bush, Maarten Cuypers, F Roggero, and Anne Randi Syversveen. Methods for quantifying the uncertainty of production forecasts: a comparative study. *Petroleum Geoscience*, 7(S):S87–S96, 2001. [cited at p. 8]

[56] Edward W Forgy. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21:768–769, 1965. [cited at p. 137]

[57] James E Fowler. Compressive-projection principal component analysis. *Image Processing, IEEE Transactions on*, 18(10):2230–2242, 2009. [cited at p. 58]

[58] Kerry Gallagher, Karl Charvin, Soren Nielsen, Malcolm Sambridge, and John Stephenson. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems. *Marine and Petroleum Geology*, 26(4):525–535, 2009. [cited at p. 136, 140, 163]

[59] GR Gavalas, PC Shah, and John H Seinfeld. Reservoir history matching by Bayesian estimation. *Old SPE Journal*, 16(6):337–350, 1976. [cited at p. 44, 52]

[60] Yulia Gel, Adrian E Raftery, and Tilmann Gneiting. Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation method. *Journal of the American Statistical Association*, 99(467):575–583, 2004. [cited at p. 136]

[61] Edward I George. The variable selection problem. *Journal of the American Statistical Association*, 95(452):1304–1308, 2000. [cited at p. 141]

[62] BOX Gep, Gailim M Jenkins, and Gregory C Reinsel. Time series analysis: forecasting and control. *Oakland CA: Holden-Dag*, 1976. [cited at p. 94]

[63] Thomas Gerstner and Michael Griebel. Numerical integration using sparse grids. *Numerical algorithms*, 18(3-4):209–232, 1998. [cited at p. 158]

[64] Roger G Ghanem and Pol D Spanos. *Stochastic finite elements: a spectral approach*. Courier Dover Publications, 2003. [cited at p. 9, 12, 13, 30, 45, 53, 63, 103]

[65] Simon J Godsill. *On the relationship between MCMC model uncertainty methods*. Citeseer, 1998. [cited at p. 140]

[66] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. The John Hopkins Press, Baltimore, 2012. [cited at p. 55]

[67] Steven N Goodman. Toward evidence-based medical statistics. 2: The Bayes factor. *Annals of internal medicine*, 130(12):1005–1013, 1999. [cited at p. 138]

[68] Peter J Green. Reversible jump Markov chain Monte Carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995. [cited at p. 136, 140]

[69] Barıs Güyagüler. *Optimization of well placement and assessment of uncertainty*. PhD thesis, stanford university, 2002. [cited at p. 8]

[70] W Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. [cited at p. 92]

[71] Simai He, Zhening Li, and Shuzhong Zhang. Approximation algorithms for homogeneous polynomial optimization with quadratic constraints. *Mathematical programming*, 125(2):353–383, 2010. [cited at p. 65]

[72] Takeyuki Hida. *Brownian motion*. Springer, 1980. [cited at p. 12]

[73] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. [cited at p. 140]

[74] Christopher H Holloman, Herbert K H Lee, and Dave M Higdon. Multiresolution genetic algorithms and Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 15(4):861–879, 2006. [cited at p. 81]

[75] SP Huang, ST Quek, and KK Phoon. Convergence study of the truncated Karhunen–Loeve expansion for simulation of stochastic processes. *International journal for numerical methods in engineering*, 52(9):1029–1043, 2001. [cited at p. 53, 54, 103]

[76] Sastry S Isukapalli. *Uncertainty analysis of transport-transformation models*. PhD thesis, Rutgers, The State University of New Jersey, 1999. [cited at p. 34, 35]

[77] Kiyosi Itô. Multiple wiener integral. *Journal of the Mathematical Society of Japan*, 3(1):157–169, 1951. [cited at p. 12]

[78] J.D. Jansen, D.R. Brouwer, G. Naevdal, and Van Kruijsdijk. Closed-loop reservoir management. *First Break*, 23(1):43–48, 2005. [cited at p. 43, 82]

[79] Svante Janson. *Gaussian Hilbert Spaces*, volume 129. Cambridge university press, 1997. [cited at p. 12]

[80] Edwin T Jaynes. Prior probabilities. *Systems Science and Cybernetics, IEEE Transactions on*, 4(3):227–241, 1968. [cited at p. 84]

[81] Edwin T Jaynes. *Probability theory: the logic of science*. Cambridge university press, 2003. [cited at p. 81]

[82] Harold Jeffreys. *Theory of probability*. Oxford University Press, 1998. [cited at p. 136]

[83] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. [cited at p. 135]

[84] William B Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984. [cited at p. 55, 57]

[85] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005. [cited at p. 70, 100]

[86] A.G. Journel. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical geology*, 34(5):573–596, 2002. [cited at p. 51]

[87] Heinrich Junker, Thies Dose, Laetitia Plas, and Adam Little. Modern approach to estimation of uncertainty of predictions with dynamic reservoir simulation-a case study of a German rotliegend gasfield. In *SPE Annual Technical Conference and Exhibition*, 2006. [cited at p. 8]

[88] Gopinath Kallianpur. *Stochastic filtering theory*. Springer, 1980. [cited at p. 12, 13]

[89] Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995. [cited at p. 136, 140, 141]

[90] Victor J Katz. The history of Stokes' theorem. *Mathematics Magazine*, 52(3):146–156, 1979. [cited at p. 4]

[91] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. Wiley-Interscience, 2009. [cited at p. 134]

[92] Maurice George Kendall. *Multivariate analysis*. Macmillan New York, 1980. [cited at p. 134]

[93] A. D. Kiureghian and O. Ditlevsen. Aleatory or epistemic? does it matter? *Structural Safety*, 31(2):105–112, 2009. [cited at p. 11]

[94] Judy L Klein. *Statistical visions in time: a history of time series analysis, 1662-1938*. Cambridge University Press, 1997. [cited at p. 138]

[95] Joseph B Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2):115–129, 1964. [cited at p. 134]

[96] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001. [cited at p. 66]

[97] Monique Laurent. Semidefinite representations for finite varieties. *Mathematical Programming*, 109(1):1–26, 2007. [cited at p. 66]

[98] Edward E Leamer. *Specification searches: ad hoc inference with nonexperimental data*. Wiley New York, 1978. [cited at p. 141]

[99] B Li and F Friedmann. Novel multiple resolutions design of experiment/response surface methodology for uncertainty analysis of reservoir simulation forecasts. In *SPE Reservoir Simulation Symposium*, 2005. [cited at p. 8]

[100] Heng Li, Pallav Sarma, and Dongxiao Zhang. A comparative study of the probabilistic-collocation and experimental-design methods for petroleum-reservoir uncertainty quantification. *SPE Journal*, 16(2):429–439, 2011. [cited at p. 36, 102, 106]

[101] Ruijian Li, AC Reynolds, and DS Oliver. History matching of three-phase flow production data. *SPE Journal*, 8(4):328–340, 2003. [cited at p. 80]

[102] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003. [cited at p. 145]

[103] N. Liu and D. Oliver. Automatic history matching of geologic facies. *SPE Journal*, 9(4):429–436, 2004. [cited at p. 43, 81]

[104] Xianlin Ma, Mishal Al-Harbi, Akhil Datta-Gupta, and Yalchin Efendiev. An efficient two-stage sampling method for uncertainty quantification in history matching geological models. *SPE Journal*, 13(1):77–87, 2008. [cited at p. 81, 82]

[105] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, page 14. California, USA, 1967. [cited at p. 137]

[106] David Madigan and Adrian E Raftery. Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546, 1994. [cited at p. 141]

[107] David Madigan, Jeremy York, and Denis Allard. Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, pages 215–232, 1995. [cited at p. 140]

[108] Paul Malliavin. *Stochastic analysis*, volume 199. Springer Berlin, 1997. [cited at p. 14]

[109] P Masani. Wiener's contributions to generalized harmonic analysis, prediction theory and filter theory. *Bulletin (New Series) of the American Mathematical Society*, 72:73–125, 1966. [cited at p. 12]

[110] Hermann G Matthies and Christian Bucher. Finite elements for stochastic media problems. *Computer methods in applied mechanics and engineering*, 168(1):3–17, 1999. [cited at p. 13, 63]

[111] Robert McCulloch and Peter E Rossi. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1):207–240, 1994. [cited at p. 140]

[112] Geoffrey J. McLachlan, RW Bean, and David Peel. A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics*, 18(3):413–422, 2002. [cited at p. 137]

[113] Dennis McLaughlin and Lloyd R Townley. A reassessment of the groundwater inverse problem. *Water Resources Research*, 32(5):1131–1161, 1996. [cited at p. 7]

[114] James Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pages 415–446, 1909. [cited at p. 62]

[115] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21:1087, 1953. [cited at p. 89, 92, 93]

[116] Nicholas Metropolis and Stanislaw Ulam. The Monte Carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949. [cited at p. 92, 93]

[117] Andrew D Miall. *The geology of fluvial deposits*, volume 575. Springer Berlin, 1996. [cited at p. 67]

[118] W.J. Miliken, A. Emanuel, and A. Chakravarty. Application of 3d streamline simulation to assist history matching. Dallas, TX, 2000. SPE paper 63155 presented at the SPE Annual Technical Conference and Exhibition. [cited at p. 43]

[119] J Milton and J Arnold. Introduction to probability and statistics: principles and applications for engineering and the computer sciences, 1995. [cited at p. 85]

[120] Douglas C Montgomery. *Design and analysis of experiments*. Wiley. com, 2006. [cited at p. 102, 104]

[121] Y. Nesterov. Squared functional systems and optimization problems. *High performance optimization*, 33:405–440, 2000. [cited at p. 66]

[122] Harald Niederreiter. Low-discrepancy and low-dispersion sequences. *J. Number Theory*, 30(1):51–70, 1988. [cited at p. 37, 38]

[123] Harald Niederreiter. *Quasi-Monte Carlo Methods*. Wiley Online Library, 1992. [cited at p. 34, 158, 159]

[124] Erich Novak and Klaus Ritter. High dimensional integration of smooth functions over cubes. *Numerische Mathematik*, 75(1):79–97, 1996. [cited at p. 158]

[125] Dean Oliver. Multiple realizations of the permeability field from well test data. *SPE Journal*, 1(2):145–154, 1996. [cited at p. 44, 52]

[126] Dean S Oliver and Yan Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15(1):185–221, 2011. [cited at p. 6, 7, 8, 43, 77, 80]

[127] Dean S Oliver, Luciane B Cunha, and Albert C Reynolds. Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology*, 29(1):61–91, 1997. [cited at p. 81]

[128] Dean S Oliver, Albert C Reynolds, Zhuoxin Bi, and Yafes Abacioglu. Integration of production data into reservoir models. *Petroleum Geoscience*, 7(S):S65–S73, 2001. [cited at p. 8]

[129] Dean S Oliver, Albert C Reynolds, and Ning Liu. *Inverse theory for petroleum reservoir characterization and history matching*. Cambridge University Press, 2008. [cited at p. 77, 82]

[130] Umut Ozdogan, Akshay Sahni, Burak Yeten, Baris Guyaguler, and Wen Chen. Efficient assessment and optimization of a deepwater asset using fixed pattern approach. In *SPE Annual Technical Conference and Exhibition*, 2005. [cited at p. 8]

[131] Yan Pan and Roland Horne. Improved methods for multivariate optimization of field development scheduling and well placement design. In *SPE Annual Technical Conference and Exhibition*, 1998. [cited at p. 8]

[132] Athanasios Papoulis and S Unnikrishna Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002. [cited at p. 19, 91, 92]

[133] Hyucksoo Park, Céline Scheidt, Darryl Fenwick, Alexandre Boucher, and Jef Caers. History matching and uncertainty quantification of facies models with multiple geological interpretations. *Computational Geosciences*, pages 1–13, 2013. [cited at p. 152]

[134] Beresford N Parlett. *The symmetric eigenvalue problem*, volume 7. SIAM, 1980. [cited at p. 58]

[135] Pablo A Parrilo. *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*. PhD thesis, Citeseer, 2000. [cited at p. 66]

[136] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894. [cited at p. 138]

[137] Cheong Peng and Ritu Gupta. Experimental design and analysis methods in multiple deterministic modelling for quantifying hydrocarbon in-place probability distribution curve. In *SPE Asia Pacific Conference on Integrated Modelling for Asset Management*, 2004. [cited at p. 8]

[138] I Petrovska and J. Carter. Estimation of distribution algorithms for history-matching. In *Proc. of the 10th European Conf. on the Mathematics of Oil Recovery. EAGE, Amsterdam*, 2006. [cited at p. 80]

[139] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm intelligence*, 1(1):33–57, 2007. [cited at p. 80, 81]

[140] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133(5):1155–1174, 2005. [cited at p. 136]

[141] Adrian E Raftery, David Madigan, and Jennifer A Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997. [cited at p. 140, 141, 146]

[142] Matthew T Reagana, Habib N Najm, Roger G Ghanem, and Omar M Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545–555, 2003. [cited at p. 36]

[143] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM review*, 26(2):195–239, 1984. [cited at p. 26]

[144] Sidney Resnick. *A probability path*. Birkhauser, 1999. [cited at p. 16]

[145] Albert C Reynolds, Nanqun He, Lifu Chu, and Dean S Oliver. Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data. *SPE Journal*, 1(4):413–426, 1996. [cited at p. 44, 52]

[146] Paul Richard and Halmos. *Introduction to Hilbert space and the theory of spectral multiplicity*. Chelsea New York, 1957. [cited at p. 17]

[147] Sylvia Richardson and Peter J Green. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792, 1997. [cited at p. 140]

[148] E Rietsch. The maximum entropy approach to inverse problems. *J. Geophys*, 42:489–506, 1977. [cited at p. 84]

[149] Gareth O Roberts and Jeffrey S Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001. [cited at p. 93, 94]

[150] Temístocles Rojas, Vasily Demyanov, Mike Christie, and Dan Arnold. Reducing uncertainty in modelling fluvial reservoirs by using intelligent geological priors. In *Geostatistical Congress, Olso*, 2012. [cited at p. 152, 153]

[151] Malcolm Sambridge. Geophysical inversion with a neighbourhood algorithmÚii. Appraising the ensemble. *Geophysical Journal International*, 138(3):727–746, 1999. [cited at p. 80, 81]

[152] P. Sarma. *Efficient closed-loop optimal control of petroleum reservoirs under uncertainty*. PhD thesis, Stanford University, 2006. [cited at p. 9, 32]

[153] P. Sarma, L. Durlofsky, K. Aziz, and W. Chen. A new approach to automatic history matching using kernel PCA. Houston,, 2007. SPE paper 106176, presented at the SPE Reservoir Simulation Symposium. [cited at p. 43]

[154] Pallav Sarma, Louis J Durlofsky, and Khalid Aziz. Kernel principal component analysis for efficient, differentiable parameterization of multipoint geostatistics. *Mathematical geosciences*, 40(1):3–32, 2008. [cited at p. 44, 53, 54, 55, 56, 57, 60, 61, 62, 63, 67]

[155] Pallav Sarma, Louis J Durlofsky, Khalid Aziz, and Wen H Chen. Efficient real-time reservoir management using adjoint-based optimal control and model updating. *Computational Geosciences*, 10(1):3–36, 2006. [cited at p. 43, 44, 45, 82]

[156] Pallav Sarma and Jiang Xie. Efficient and robust uncertainty quantification in reservoir simulation with polynomial chaos expansions and non-intrusive spectral projection. In *Reservoir Simulation Symposium*, 2011. [cited at p. 36, 37]

[157] C Scheidt and J Caers. A new method for uncertainty quantification using distances and kernel methods. Application to a deepwater turbidite reservoir. *SPEJ, SPE-118740-PA*, 2009. [cited at p. 152]

[158] Celine Scheidt and Jef Caers. Uncertainty quantification in reservoir performance using distances and kernel methods–application to a West Africa deepwater turbidite reservoir. *SPE Journal*, 14(4):680–692, 2009. [cited at p. 152]

[159] Schlumberger. Keeping pace with oilfield complexity. *Oilfield Review*, 23, 2011. [cited at p. iii, 6, 7]

[160] Petrel Schlumberger. Enhance precision geosteering and decision making with real-time 3d modeling using Petrel workflow tools, 2005. [cited at p. 69]

[161] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998. [cited at p. 53, 55, 56, 60, 61, 62, 63]

[162] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978. [cited at p. 136]

[163] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978. [cited at p. 140]

[164] Nati Shalom. The scalability revolution: From dead end to open road. *GigaSpaces Concept Paper*, 2007. [cited at p. 6]

[165] NZ Shor. Quadratic optimization problems. *Soviet Journal of Circuits and Systems Sciences*, 25(1-11):6, 1987. [cited at p. 66]

[166] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. Chapman & Hall/CRC, 1986. [cited at p. 26]

[167] Ian H Sloan and Henryk Woźniakowski. When are quasi-monte carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33, 1998. [cited at p. 158]

[168] Per Arne Slotte and Eivind Smorgrav. Response surface methodology approach for history matching and uncertainty assessment of reservoir simulation models (SPE-113390). In *70th EAGE Conference & Exhibition*, 2008. [cited at p. 8]

[169] Ilya M Sobol. On quasi-monte carlo integrations. *Mathematics and Computers in Simulation*, 47(2):103–112, 1998. [cited at p. 158]

[170] Elias M Stein and Rami Shakarchi. *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, 2009. [cited at p. 16]

[171] HL Stone. Probability model for estimating three-phase relative permeability. *Journal of Petroleum Technology*, 22(2):214–218, 1970. [cited at p. 5]

[172] O Stramer and RL Tweedie. Langevin-type models ii: Self-targeting candidates for MCMC algorithms*. *Methodology and Computing in Applied Probability*, 1(3):307–328, 1999. [cited at p. 81]

[173] S. Strebelle and Sebastien Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002. [cited at p. 51, 54]

[174] Sebastien Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002. [cited at p. 153]

[175] Markus Svensén and Christopher M Bishop. Robust Bayesian mixture modelling. *Neurocomputing*, 64:235–252, 2005. [cited at p. 135, 136]

[176] Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005. [cited at p. 65, 81, 83, 84, 85, 87, 110]

[177] Menner A Tatang. *Direct incorporation of uncertainty in chemical and environmental engineering systems*. PhD thesis, Massachusetts Institute of Technology, 1995. [cited at p. 32]

[178] Kardi Teknomo. K-means clustering tutorial. *Medicine*, 100(4):3, 2006. [cited at p. 135]

[179] Robert Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996. [cited at p. 140]

[180] Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986. [cited at p. 140]

[181] Radu Alexandru Todor and Christoph Schwab. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA Journal of Numerical Analysis*, 27(2):232–261, 2007. [cited at p. 13, 63]

[182] Simon Tong and Daphne Koller. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research*, 2:45–66, 2002. [cited at p. 152]

[183] John A Trangenstein and John B Bell. Mathematical structure of the black-oil model for petroleum reservoir simulation. *SIAM Journal on Applied Mathematics*, 49(3):749–783, 1989. [cited at p. 5]

[184] A. W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. [cited at p. 16]

[185] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996. [cited at p. 66]

[186] Jasper A Vrugt, CJF Ter Braak, CGH Diks, Bruce A Robinson, James M Hyman, and Dave Higdon. Accelerating Markov chain Monte carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3):273–290, 2009. [cited at p. 81]

[187] Xiaoliang Wan and George Em Karniadakis. Beyond Wiener–Askey expansions: handling arbitrary pdfs. *Journal of Scientific Computing*, 27(1-3):455–464, 2006. [cited at p. 39, 134]

[188] Xiaoliang Wan and George Em Karniadakis. Multi-element generalized polynomial chaos for arbitrary probability measures. *SIAM Journal on Scientific Computing*, 28(3):901–928, 2006. [cited at p. 39]

[189] Larry Wasserman. Asymptotic inference for mixture models by using data-dependent priors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):159–180, 2000. [cited at p. 140]

[190] AT Watson, JG Wade, and RE Ewing. Parameter and system identification for fluid flow in underground reservoirs. In *Proceedings of the Conference, Inverse Problems and Optimal Design in Industry, Philadelphia, PA*. Citeseer, 1994. [cited at p. 7]

[191] JW Watts. 38441-reservoir simulation: Past, present, and future. *SPE Computer Applications*, 9(6):171–176, 1997. [cited at p. 6]

[192] M. Webster and M.A. Tatang. Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model. *technical report from the MIT Joint Program on the Science and Policy of Global Change, Cambridge, MA,*, 1996. [cited at p. 32, 63]

[193] Walter Frank Raphael Weldon. Mendel's laws of alternative inheritance in peas. *Biometrika*, 1(2):228–254, 1902. [cited at p. 138]

[194] Norbert Wiener. Differential space. *Journal of Mathematical Physics*, 2:131–174, 1923. [cited at p. 11]

[195] Norbert Wiener. Generalized harmonic analysis. *Acta Mathematica*, 55(1):117–258, 1930. [cited at p. 11, 12]

[196] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938. [cited at p. 11, 17]

[197] Z. Wu. A Newton-Raphson iterative scheme for integrating multiphase production data into reservoir models. *SPE Journal*, 6:343–351, 2001. [cited at p. 43, 81]

[198] Dongbin Xiu and Jan S Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139, 2005. [cited at p. 13, 63]

[199] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002. [cited at p. ii, x, 29, 38, 39, 63, 133]

[200] William WG Yeh. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, 22(2):95–108, 1986. [cited at p. 7]

[201] B. Yeten. Optimization of field development. In *Proceedings of 9th International Forum on Reservoir Simulation, Abu Dhabi, UAE, 9-13 December*, 2007. [cited at p. 8]

[202] Fengjun Zhang and Albert C Reynolds. Optimization algorithms for automatic history matching of production data. In *8th European Conference on the Mathematics of Oil Recovery*, 2002. [cited at p. 80]

[203] Denis Zubarev. Pros and cons of applying proxy-models as a substitute for full reservoir simulations. In *SPE Annual Technical Conference and Exhibition*, 2009. [cited at p. 9]